Feature Identification of False Health Information on Social Media (Postprint)

Authors: Zhang Shuai

Date: 2023-04-01T16:02:50+00:00

Abstract

[Purpose/Significance] This study aims to identify the characteristics of false health information on social media and construct a feature checklist for such information, thereby providing theoretical support for the measurement of these characteristics and offering a valuable reference for both users and social media platforms in discerning false health information. [Method/Process] A total of 1,004 social media health data entries were collected. Programmatic coding was employed to extract key features of false health information on social media. Chi-square tests and analysis of variance were utilized to reveal the significant characteristics of false health information on social media, and a feature checklist for social media false health information was constructed. [Results/Conclusion] The research findings indicate that the characteristics of false health information on social media comprise three dimensions—surface features, semantic features, and source features—encompassing 11 primary characteristics and 29 sub-characteristics. Specifically, false health information on food safety themes in social media exhibits more prominent "terminology packaging" characteristics; "fact exaggeration" constitutes a significant characteristic of false health information on common disease themes in social media; and false health information on health preservation and wellness themes in social media is characterized by significant "metadata absence" and "authority appropriation" features.

Full Text

Preamble

Social media platforms allow any user to directly publish and disseminate health information without verifying its authenticity or reliability [21]. Existing research has examined health information from the perspectives of authority, traceability, rationality, transparency, advertising, and editorial policy [32].

LIDA, a tool developed by an Oxford University-affiliated company for identifying false healthcare information on websites, consists of three components: accessibility, usability, and reliability [33]. JAMA, a set of standards developed by the Journal of the American Medical Association for identifying false health information on medical websites, comprises four dimensions: author, attribution, disclosure, and currency, with specific evaluation indicators including author name, affiliation, credentials, references, sources, conflicts of interest, advertising, and dates [34]. Although these assessment tools have been developed, most focus on false information in medical and health websites, and their applicability to social media contexts remains to be verified. Furthermore, studies show that users' motivation to share personal health experiences on social media is often stronger than on other online platforms, yet this first-hand health information may be inaccurate [22]. Compared to health information on websites, social media health information tends to be simplified, potentially omitting small but important details [21]. Given these considerations, this study focuses on false health information on social media, where health information is defined as all knowledge, techniques, skills, concepts, and behaviors related to human health [23]. Following L. Bode and E.K. Vraga's perspective [24], false health information is defined as erroneous health information lacking scientific evidence and expert support. This definition can describe various types of erroneous health information on social media, such as rumors, distorted health information, pseudo-health information, and other variants [25].

2.2 Research on Characteristics of False Information

As false information proliferates online, scholars have investigated its characteristics. J. Zhou et al. examined health rumors on Twitter and identified seven features: emotional valence, appeal, publisher authority, external evidence, argument length, hashtags, and direct messages [26]. Y. Li et al. used the CARS checklist to distill online false health information features into four dimensions: lack of credibility, lack of accuracy, lack of rationality, and lack of relevant support [27]. L. Rubin categorized online false information features into three aspects: fabrication, concealment (omission of major facts), and ambiguity [28]. L. Zhou et al. classified website false information features into nine categories: quantity, complexity, uncertainty, indirectness, expressiveness, diversity, informality, specificity, and influence [29]. L. Lavoragna et al. analyzed characteristics of false health news in online virtual communities, finding that false health news often exaggerates scientifically unverified facts [30]. While these studies have identified some features of false information, most focus on foreign-language contexts and require validation for Chinese-language applicability. Moreover, feature extraction has rarely been based on empirical data, and research attention to false health information characteristics on social media remains limited.

2.3 Research on False Information Identification

Current studies have attempted to develop evaluation tools to identify false information. DISCERN, developed by the Oxford University Health Sciences Research Institute, is a widely used tool to help consumers identify false treatment information on the internet, containing 15 key indicators related to clarity, relevance, appropriateness, references, dates, and objectivity of treatment information [31]. HONcode, established by the Swiss Health On the Net Foundation, is a set of guidelines to help patients, healthcare workers, and the general public identify false information on health websites, comprising eight evaluation principles [32]. Some scholars have also employed machine learning methods to identify false information. For instance, M. Ott et al. [35], S. Shojaeie et al. [36], and J. Li et al. [37] used bag-of-words, part-of-speech, and stylistic features to extract text from manually compiled false and authentic reviews, constructing Naive Bayes (NBM) and Support Vector Machine (SVM) models that achieved 84%-89.6% identification accuracy. N. Jindal and B. Liu [38-39] used logistic regression models on Amazon datasets based on stylistic, metadata, and grammatical features, achieving 63%-78% accuracy. A. Mukherjee et al. [40] employed SVM classifiers on Yelp datasets using textual features, achieving 65.6% and 67.8% accuracy. While these methods using neural networks, decision trees, and logistic regression have achieved certain effects, most current research focuses only on partial features of false information while neglecting other aspects. Therefore, this study collects real social media health data and employs a mixed-method approach combining content analysis and statistical analysis to comprehensively investigate the overall characteristics of false health information on social media.

3 Research Design and Implementation

3.1 Research Methods and Tools

This study adopts a non-interventional research method—content analysis—using NVivo Plus software to extract features of false health information on social media. Content analysis is a scientific method for inductively and deductively analyzing raw materials, enabling objective and reliable identification of core themes through fine-grained analysis of original data [41]. NVivo Plus is a mainstream tool for data coding analysis, and its NCapture plugin has functions for collecting and analyzing social media data, particularly suitable for text data acquisition and analysis [42]. Based on extracted features of false health information on social media, this study further employs chi-square tests and analysis of variance to reveal significant characteristics of false health information.

3.2 Sample Selection

This study selected false health information disseminated on WeChat as the research sample. As a representative mainstream social media platform, WeChat provides users with abundant health information and related services, making it one of China's most influential social media platforms. Research has found that over half of health information on WeChat is false [10], making it a representative and typical platform for this study. This study utilized the WeChat Rumor Debunking Assistant to screen false health information. This mini-program, developed by WeChat, has been joined by over 400 authoritative institutions, government organizations, and mainstream media outlets (such as People's Daily Online, Voice of Chinese Academy of Sciences, Science Popularization China, Dingxiang Doctor, and Guokr). Its primary purpose is to expose false information lacking scientific basis [43]. The assistant also publishes scientifically verified authentic health information to refute corresponding false information. Using NCapture, the researcher collected 502 pieces of false health information on WeChat confirmed as false between January 2018 and August 2020. During the same period, 502 pieces of authentic health information refuting these false claims were also collected. Additionally, health information whose authenticity could not be determined was excluded. The researcher then reviewed all 1,004 pieces of true and false health information, saved them as PDF files, and established a data index for each piece. The sample collection lasted two months.

Based on the World Health Organization's (WHO) classification of health topics [44], two researchers jointly categorized the theme of each piece of true and false health information. The data were ultimately classified into three themes: (1) Food safety (31.3%), including food additives, genetically modified foods, and food combinations; (2) Common diseases (27.7%), including influenza, vaccines, hypertension, diabetes, and heart disease; and (3) Health and wellness (41.0%), including health regimens, fitness and weight loss, and maternal care. Since authentic and false health information were paired, they shared the same themes, with each theme containing equal numbers of true and false messages (50% each).

3.3 Data Coding

Based on the International Federation of Library Associations and Institutions (IFLA) 2016 information chart "How To Spot Fake News" [45], this study constructed a coding framework for false health information features on social media. The IFLA chart was developed to help the general public identify false information in media, considering information authenticity from eight aspects: considering news sources, reading in full, checking author information, examining arguments, verifying dates, checking if it's a joke, checking personal biases, and consulting experts. To ensure coding accuracy, the researcher adapted the IFLA framework to the research context, appropriately summarizing features from the chart. "Checking author information" and "verifying dates" revealed the "metadata deficiency" feature; "considering news sources" revealed the "source ambiguity" feature; "reading in full" revealed the "incomplete information" feature; "arguments" revealed the "lack of verification" feature; and "is it a joke" revealed the "exaggerated facts" feature. The features "checking personal biases" and "consulting experts" were not directed at the false information itself and were therefore excluded. This resulted in five key features: source ambiguity, incomplete information, metadata deficiency, lack of verification, and exaggerated facts, which formed the initial coding framework (see Table 1).

Since the IFLA chart was developed in an English-language context and may not be fully applicable to Chinese contexts, and since this study focuses on health information on social media rather than general media information, the coding process was not limited to existing features in the framework. To ensure scientific rigor, the researcher followed J.M. Corbin and A. Strauss's procedural coding method [46], conducting open coding, axial coding, and selective coding. Since this study required comparative analysis between false and authentic health information features to identify key characteristics of false health information on social media, authentic health information was analyzed using the same coding method, with features coded separately.

Two coders performed the coding. To ensure reliability, 20% of the data were randomly selected for dual coding, yielding a Cohen's kappa coefficient of 0.82, indicating high inter-coder consistency [47]. The remaining 80% of data were randomly divided into two groups for separate coding. Disagreements during coding were resolved through group discussions to select the feature most relevant to the research topic.

Phase I (Code I) Open Coding: Researchers reviewed each piece of health information line by line in NVivo Plus. Through careful examination of raw materials and repeated induction, with slight vocabulary standardization, 29 initial categories and 4,478 coding reference points were extracted, including unclear sources, incomplete content, missing author information, lack of scientific basis, exaggerated efficacy, immoderate tone, impersonation of authoritative media, terminology abuse, spelling errors, claims of breaking news, and fabricated cases.

Phase II (Code II) Axial Coding: Researchers mapped the 29 initial categories to the false health information feature coding framework based on logical relationships between subcategories and main categories, ultimately forming 11 main categories. These included features from the coding framework (source ambiguity, incomplete information, metadata deficiency, lack of verification, exaggerated facts) and newly developed features beyond the framework (improper tone/language, false authority, terminology packaging, format confusion, information inducement, and fabricated information).

Phase III (Code III) Selective Coding: Researchers used meaning construction to integrate and focus the main categories [48], merging the 11 main categories into three core categories: surface features, semantic features, and source features. Surface features refer to information format correctness, including metadata deficiency and format confusion. Semantic features refer to information content accuracy, including incomplete information, exaggerated facts, improper tone/language, terminology packaging, information inducement, and fabricated information. Source features refer to information source credibility, including source ambiguity, lack of verification, and false authority.

The coding visualization is shown in Figure 1 [Figure 1: see original paper], where different branch colors represent coding for different false health information features, branch width represents coding proportion (source features 31.2%, semantic features 49.6%, surface features 19.2%), flow direction represents relationships between coding nodes, and adjacent vertical line areas represent coding phases (Code I, Code II, Code III). This figure intuitively reflects the coding process and relationships of false health information features on social media.

3.4 Theoretical Saturation Test

After three-level coding, the features of false health information on social media were basically determined. To ensure theoretical saturation, researchers randomly selected 30 new pieces of false health information 10 consecutive times for coding using the same rules and procedures [49]. No new coding categories or relationships emerged, indicating that the features of false health information on social media had been fully captured and theoretical saturation was achieved.

4 Data Analysis

Through data coding, this study extracted false health information features from social media with 3 dimensions, 11 main features, and 29 sub-features. To more precisely identify significant features of false health information on social media, the researcher 统计了真假健康信息在不同主题和不同特征维度上的编码分布, with data summarized in Table 2 . In Table 2, F represents false health information, T represents authentic health information, values outside parentheses represent frequencies of false features, and values inside parentheses represent proportions of false feature occurrences.

Based on this, chi-square tests were used to identify significant features of false health information within the same theme, and analysis of variance was used to compare significant features across different themes.

4.1 Analysis of Significant Features of False Health Information Within the Same Theme

Chi-square tests were conducted for each feature within the same theme (see Table 3) to test whether significant differences existed between true and false health information features for the same social media health topic. In this study, when the approximate probability value (p-value) corresponding to the ^{2} statistic was less than the significance level (α=0.05), true and false health information features were considered significantly different. Results showed that for food safety, common diseases, and health and wellness themes, significant differences existed between true and false health information features. Surface features, semantic features, and source features, along with their sub-features, could effectively distinguish false from authentic health information. If social media health information exhibited prominent characteristics across these feature dimensions, it could be identified as false health information.

4.2 Analysis of Significant Features of False Health Information Across Different Themes

Analysis of variance was used to test whether significant differences existed in false health information features across different themes. Results (see Table 4) showed that features including “metadata deficiency,” “exaggerated facts,” “improper tone/language,” “terminology packaging,” “information inducement,” “fabricated information,” “source ambiguity,” “lack of verification,” and “false authority” exhibited significant differences across different themes. Post-hoc comparisons revealed that:

- The food safety theme (Group 1) showed more significant “terminology packaging” features ($p_{1,2}$=.000**, $p_{1,3}$=.011) compared to other themes, frequently using professional terms or pseudo-concepts like “hormone-induced growth technology,” “azodicarbonamide,” and “aspartame” to create a false appearance of professionalism that confuses users.
- The common diseases theme (Group 2) showed more significant “exaggerated facts” features ($p_{1,2}$=.000**, $p_{2,3}$=.039) compared to other themes, commonly using phrases like “shocking the world,” “number one miracle cure,” “anti-cancer star,” and “radical cure” to claim extremely high cure rates.
- The health and wellness theme (Group 3) showed more significant “metadata deficiency” ($p_{1,3}$=.000, $p_{2,3}$=.043) **and “false authority”** ($p_{1,3}$=.000*, $p_{2,3}$=.008**) features compared to other themes, often lacking important information such as “author name,” “publication date,” “reviewer,” and “copyright statement,” while frequently using titles like “Nobel Prize winner,” “chief scientist,” and “academician” to conduct false advertising for health products.

4.3 Construction of a Feature Checklist for False Health Information on Social Media

Based on chi-square tests and analysis of variance, this study identified significant features of false health information on social media. Building on these significant features, the researcher constructed a feature checklist for false health information on social media to provide users with useful references for better identification (see Table 5). The checklist primarily consists of feature dimensions, main features, sub-features, and related recommendations for false health information on social media. Users can apply these key features and recommendations to comprehensively evaluate health information on social media.

5 Research Conclusions and Implications

This study collected 1,004 pieces of empirical data from social media networks, extracted 11 main features of false health information using content analysis, and integrated and clustered them into three dimensions: surface features, semantic features, and source features. Among these, "metadata deficiency," "exaggerated facts," "information inducement," "incomplete information," "improper tone/language," "source ambiguity," "lack of verification," and "false authority" were important features of false health information on social media, while "format confusion," "fabricated data," and "terminology packaging" were secondary features.

Building on this, chi-square tests identified surface features, semantic features, and source features as significant characteristics for distinguishing false from authentic health information within the same theme. Analysis of variance identified "metadata deficiency," "exaggerated facts," "terminology packaging," and "false authority" as significant features across different themes. Specifically, food safety-related false health information on social media exhibited more prominent "terminology packaging"; "exaggerated facts" were significant features for common disease-related false health information; and health and wellness-related false health information showed significant "metadata deficiency" and "false authority" features. These findings will facilitate automatic identification of theme-specific false health information. Ultimately, this study constructed a feature checklist for false health information on social media, which not only provides theoretical support for measuring false health information features but also enables social media platforms to build automatic filtering mechanisms for false health information.

Compared with existing research [6, 51-52], this study makes three theoretical contributions: (1) While previous studies mostly used interventional methods such as interviews, expert consultation, and surveys to evaluate online health information features—methods that 介入研究对象的活动 to varying degrees and make it difficult to objectively reflect authentic evaluation results—this study employed non-interventional methods to scientifically analyze authentic social media health data, forming a more reliable feature checklist; (2) This study focused on grammatical, semantic, and pragmatic features of false health information on social media, providing an informatics research perspective for measuring false health information features; (3) This study explored features of theme-specific false health information in social media environments, identifying "metadata deficiency," "exaggerated facts," "terminology packaging," and "false authority" as significant features for different themes, which has received little attention in previous research.

In practical terms, this study provides not only an effective tool for social media users to identify false health information but also a viable approach for social media service providers to discriminate false health information. For social media users, the naming of false health information features in this study originates from relevant original statements, intuitively revealing characteristics of false health information. Users can actively identify false health information they encounter by referring to the feature checklist, reducing the risk of serious medical consequences from believing false information. Additionally, users can improve their health information literacy and identification abilities by learning and mastering these features. For social media service providers, the feature checklist constructed in this study provides empirical support for platform managers to screen and remove false health information and theoretical support for building automatic early warning and filtering mechanisms based on this framework, thereby curbing the spread of false health information at its source. Meanwhile, publishers of health information on social media can also use this checklist to improve the surface, semantic, and source features of the health information they disseminate, thereby comprehensively improving the quality of health information on social media platforms.

Despite these contributions, this study has several limitations. First, the feature checklist for false health information on social media is still in a preliminary exploration stage. Given the dynamic nature of social media health information, the checklist should be continuously improved and optimized in future practical applications. Second, this study was based on the WeChat platform, and whether its conclusions apply to other social media platforms requires further comparative research. Finally, this study only focused on feature identification of false health information on social media without considering identification effectiveness. Future research will consider applying machine learning to social media health information datasets based on these false information features to more accurately identify false health information on social media, thereby deepening and extending the findings of this study.

Acknowledgments: Thanks to Professor Ma Feicheng for his guidance and assistance in topic selection, writing, and finalizing this paper.

References

[1] Jin Xiaoling, Feng Huihui, Zhou Zhongyun. Research on health information dissemination behavior in WeChat Moments [J]. Management Science, 2017, 30(1): 73-82. [2] Xu Xiaoting, Zhao Yuxiang, Zhu Qinghua. An empirical study on health information needs of elderly users in online health communities [J]. Library and Information Service, 2019, 63(10): 87-96. [3] Zhang X, Wen D, Liang J, et al. How the public uses social media WeChat to obtain health information in China: a survey study [J]. BMC medical informatics and decision making, 2017, 17(2): 71-79. [4] Chou WYS, Hunt YM, Beckjord EB, et al. Social media use in the United States: implications for health communication [J]. Journal of medical internet research, 2009, 11(4): e48. [5] Hou Xiaorong, Fu Yang, Chen Juan. Research on user perception and effectiveness of health information based on WeChat platform [J]. Modern Intelligence, 2016, 36(10): 89-93. [6] Li Yuelin, Zhang Xiu, Wang Shanshan. Research on quality of health information in social media: based on analysis

of true and false health information features [J]. Journal of the China Society for Scientific and Technical Information, 2018(3): 294-304. [7] Mo Zuying, Ma Feicheng. Game analysis of information resource quality control in network environment [J]. Information Studies: Theory & Application, 2012, 35(8): 26-30. [8] Goldacre B. Media misinformation and health behaviors [J]. The lancet oncology, 2009, 10(9): 848. [9] China Internet Network Information Center. The 44th Statistical Report on China's Internet Development [EB/OL]. [2020-08-21]. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201908/P020190830356787490958.htm. [10] Li Yuelin, Zhang Xiu. Research on college students' ability to identify health information in social media [J]. Library and Information Knowledge, 2018(1): 66-77. [11] Pew Research Center. Mobile technology and its social impact survey 2018 [EB/OL]. [2020-08-21]. https://www.pewresearch.org. [12] Zhang Z, Zhang Z, Li H. Predictors of the authenticity of internet health rumors [J]. Aslib journal of information management, 2017, 69(6): 660-673. [13] Diviani N, Putte B, Giani S, et al. Low health literacy and evaluation of online health information: a systematic review of the literature [J]. Journal of medical internet research, 2015, 17(5): e112. [14] Song Shijie, Zhao Yuxiang, Song Xiaokang, et al. Research on factors influencing credibility judgment of distorted health information in internet environment [J]. Journal of Library Science in China, 2019, 45(4): 72-85. [15] Yuan Jianxia, Leng Fuhai. Intelligence investigation and analysis in "infodemic" response [J]. Competitive Intelligence, 2020, 16(3): 9-13. [16] Wang Shiwei. On ten characteristics of "infodemic" [J]. Library Journal, 2020, 39(3): 19-23. [17] Devine T, Broderick J, Harris LM, et al. Making quality health websites a national public health priority: toward quality standards [J]. Journal of medical internet research, 2016, 18(8): e211. [18] Deng Shengli, Wang Fenfen. Research progress on identification of online false review information from internet governance perspective [J]. Journal of Information Resources Management, 2019(3): 73-81. [19] Wu Jiafen, Ma Feicheng. Review of research on identification methods for product false review text [J]. Data Analysis and Knowledge Discovery, 2019, 3(9): 1-15. [20] Paik A, Chua AYK, Goh DHL. Does KFC sell rat? analysis of tweets in the wake of a rumor outbreak [J]. Aslib journal of information management, 2017, 69(6): 660-673. [21] Brady JT, Kelly ME, Steins L. The trump effect: with no peer review, how do we know what to really believe on social media? [J]. Clinics in colon and rectal surgery, 2017, 30(4): 270-276. [22] Fichman RG, Kohli R, Krishnan R. The role of information systems in healthcare: current research and future trends [J]. Information systems research, 2011, 22(3): 419-428. [23] Tsai H, Bagozzi RP. Contribution behavior in virtual communities: cognitive, emotional, and social influences [J]. Management information systems quarterly, 2014, 38(1): 143-164. [24] Bode L, Vraga EK. In related news, that was wrong: the correction of misinformation through related stories functionality in social media [J]. Journal of communication, 2015, 65(4): 619-638. [25] Tandoc JRE, Lim ZW, Ling R. Defining "fake news": a typology of scholarly definitions [J]. Digital journalism, 2018, 6(2): 137-153. [26] Zhou J, Liu F, Zhou H. Understanding health food messages on Twitter for health literacy promotion [J]. Perspectives in public health, 2018, 138(3): 173-179. [27] Li Y, Zhang X, Wang S. Fake vs. real health information in social media in China [J]. Proceedings of the Association for Information Science and Technology, 2017, 54(1): 742-743. [28] Rubin VL. On deception and deception detection: content analysis of computer-mediated stated beliefs [J]. Proceedings of the Association for Information Science and Technology, 2011, 47(1): 1-10. [29] Zhou L, Burgoon JK, Nunamaker JF, et al. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications [J]. Group decision and negotiation, 2004, 13(1): 81-106. [30] Lavoragna L, De Stefano M, Sparaco M, et al. Fake news, influencers and health-related professional participation on the Web: a pilot study on a social-network of people with Multiple Sclerosis [J]. Multiple sclerosis and related disorders, 2018, 25(10): 175-178. [31] Dueñas-Garcia OF, Kandasai P, Flynn MK, et al. Patient-focused websites related to stress urinary incontinence and pelvic organ prolapse: a DISCERN quality analysis [J]. International urogynecology journal, 2015, 26(6): 875-880. [32] Boyer C, Dolamic L. Automated detection of HONcode website conformity compared to manual detection: an evaluation [J]. Journal of medical internet research, 2015, 17(6): e135. [33] Tavare AN, Alsafi A, Hamady MS. Analysis of the quality of information obtained about uterine artery embolization from the internet [J]. Cardiovascular and interventional radiology, 2012, 35(6): 1355-1362. [34] Winker MA, Flanagin A, Chi-Lum B, et al. Guidelines for medical and health information sites on the internet: principles governing AMA Websites [J]. Jama, 2000, 283(12): 1600-1606. [35] Ott M, Choi Y, Cardie C, et al. Finding deceptive opinion spam by any stretch of the imagination [C]//Proceedings of the 49th annual meeting of the NAACL-HLT. Portland: association for computational linguistics, 2011: 309-319. [36] Shojaeie S, Murad MAA, Binazlan MA, et al. Detecting deceptive reviews using lexical and syntactic features [C]//International conference on intelligent systems design and applications. Selangor, Malaysia: IEEE, 2013: 53-58. [37] Li J, Ott M, Cardie C, et al. Towards a general rule for identifying deceptive opinion spam [C]//Meeting of the Association for Computational Linguistics. Baltimore: Association for Computational Linguistics, 2014: 1566-1576. [38] Jindal N, Liu B, Lim EP. Finding unusual review patterns using unexpected rules [C]//Proceedings of the 19th ACM international conference on information and knowledge management. Toronto: ACM, 2010: 1549-1552. [39] Jindal N, Liu B. Opinion spam and analysis [C]//Proceedings of the 2008 international conference on web search and data mining. Stanford: ACM, 2008: 219-230. [40] Mukherjee A, Venkataraman V, Liu B, et al. What yelp fake review filter might be doing? [C]//Proceedings of the seventh international AAAI conference on weblogs and social media. elle vue, Washington: AAAI Press, 2013: 1-10. [41] Hsieh HF, Shannon SE. Three approaches to qualitative content analysis [J]. Qualitative health research, 2005, 15(9): 1277-1288. [42] QSR International Pty Ltd. What is nvivo [EB/OL]. [2020-08-26]. https://www.qsrinternational.com/nvivo/what-is-nvivo. [43] Liu Benying. Research on rumor propagation and immune dynamics in WeChat based on infectious disease models [D]. Wenzhou: Wenzhou University, 2019. [44] World Health Organization. Health-topics

[EB/OL]. [2020-08-26]. https://www.who.int/health-topics/. [45] International Federation of Library Associations and Institutions. How to spot fake news [EB/OL]. [2020-08-26]. https://www.ifla.org/node/11175. [46] Corbin JM, Strauss A. Grounded theory research: procedures, canons, and evaluative criteria [J]. Qualitative sociology, 1990, 13(1): 3-21. [47] Landis JR, Koch GG. The measurement of observer agreement for categorical data [J]. Biometrics, 1977, 33(1): 159-174. [48] Flick U. An introduction to qualitative research [M]. Sun Jin, trans. Chongqing: Chongqing University Press, 2011: 248-256. [49] Francis JJ, Johnston M, Robertson C, et al. What is an adequate sample size? operationalising data saturation for theory-based interview studies [J]. Psychology and health, 2010, 25(10): 1229-1245. [50] Zha X, Yang H, Yan Y, et al. Exploring the effect of social media information quality, source credibility and reputation on information fit-to-task: moderating role of focused immersion [J]. Computers in human behavior, 2018, 79(2): 227-237. [51] Lv Yalan, Hou Xiaorong, Huang Cheng, et al. Research on evaluation index system of public online health information credibility in ubiquitous network environment [J]. Journal of Intelligence, 2016, 35(1): 196-200. [52] Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? qualitative study using focus groups, usability tests, and in-depth interviews [J]. British medical journal, 2002, 324(7337): 573-577.

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