User Role Formation and Transformation on Social Q&A Platforms During Emerging Infectious Disease Outbreaks: A Case Study of Zhihu (Postprint)

Authors: Chen Miaomiao, An Lu

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] This study investigates methods for classifying user roles, key factors influencing role formation, and the characteristics and differences of role transitions on question-and-answer (Q&A) platforms in the context of emerging infectious disease outbreaks. [Methods/Procedures] A total of 702,927 Covid-19 pandemic-related data entries were collected from Q&A platforms. User roles were identified from the dimensions of participation level and value contribution. Influencing factors in community user role formation were identified based on information human factors, information factors, and information environment factors. Key factors influencing the formation of different roles were analyzed by combining multi-classification models with the SHapley Additive exPlanations (SHAP) model. Behavior patterns and thematic characteristics during different role transitions were mined using the FP-growth association rule algorithm. [Results/Conclusions] The findings indicate that users tend to maintain their roles unchanged, with transition directions primarily being active-type and lurker-type. Information volume is a key factor in the formation of different roles. The degree of change in user role transition characteristics varies significantly across different transition stages, and user role transition behaviors across all transition stages show significant differences.

Full Text

Preamble

Vol. 66 No. 12, June 2022
ChinaXiv Cooperative Journal

User Role Formation and Transformation on Social Q&A Platforms in the Context of Infectious Disease Outbreaks: A Case Study of the Zhihu Platform

Chen Miaomiao[1], An Lu[1,2]

[1] School of Information Management, Wuhan University, Wuhan 430072

[2] Center for Studies of Information Resources, Wuhan University, Wuhan 430072

Abstract: [Purpose/Significance] This study investigates user role classification methods, key factors in role formation, and the characteristics and differences of role transformation on Q&A platforms during infectious disease outbreaks. [Method/Process] A total of 702,927 Covid-19-related data entries were collected from Q&A platforms. User roles were identified from the dimensions of participation and value. Based on information user factors, information factors, and information environment factors, the influencing factors of community user role formation were identified. A multi-classification model combined with SHapley Additive exPlanations (SHAP) was used to analyze the key factors affecting different role formations, and the FP-growth association rule algorithm was employed to mine behavioral patterns and topic characteristics under different role transformations. [Result/Conclusion] The results show that users tend to maintain their roles unchanged, with transformation directions primarily toward active and lurking roles. Information volume is the key factor in the formation of different roles, and there are significant differences in the degree of change in user role transformation characteristics across different transformation stages and in user role transformation behaviors across all transformation stages.

Keywords: user role; Zhihu Q&A platform; role transformation; influencing factors; Covid-19; infectious disease outbreak

Classification Number: G252

DOI: 10.13266/j.issn.0252-3116.2022.12.007

The sudden outbreak of Covid-19 in early 2020 not only endangered public health but also brought unprecedented attention and utilization to social media platforms. Under quarantine conditions, the public used social media and other internet tools for comprehensive information searching and timely communication. However, the uncertainty of public health emergencies often caused public panic and anxiety, leading them to seek help and information satisfaction on online Q&A platforms. In response to the epidemic, General Secretary Xi Jinping emphasized the need to "control the overall public opinion and strive to create a favorable public opinion environment. We must strengthen the control of online media, promote the implementation of primary, supervisory, and regulatory responsibilities, and legally crack down on those who take the opportunity to create rumors and cause trouble" [1]. In this context, users have become important participants in social public opinion, and both community managers and government departments should strengthen their understanding of users in emergency situations, including their roles, behavioral characteristics, and transformation differences, to help relevant departments formulate targeted user policies and promote a good information exchange environment.

Zhihu is a type of social Q&A platform and currently the largest knowledge Q&A platform in China, with 100 million monthly active users in 2021. On this platform, users can browse, search, follow, ask questions, and answer them to meet their information needs or exchange information. However, most existing research on user roles focuses on communities on social platforms such as Weibo and Twitter. Different platforms have differences in user roles, behavioral characteristics, and specific research design due to their design models and positioning. Therefore, analyzing user groups on the Zhihu platform in the context of Covid-19 can supplement relevant research on virtual community user roles. Previous research has mainly focused on user role identification and classification from user behavior or user network structure [2], summarizing community user behavior patterns including participation and interaction behaviors, and identifying classic roles such as lurkers and contributors. Due to the particularity of question and answer behaviors in social Q&A communities, user roles are typically divided into lurkers, questioners, and answerers [3], with contribution levels increasing sequentially, attracting scholars to study the motivations behind user role changes, primarily using questionnaires and interviews.

In recent years, some studies have begun to focus on role transformation. On one hand, they have discovered role transfer methods. For example, J. Preece et al. proposed that users transfer from one role to another through linear or non-linear methods [9], but lacked experimental support. C. Fu proposed using a time-aware role model to effectively track user role evolution [10], abstracting user roles simply as questioners and answerers and focusing mainly on model construction. A. Bartal et al. identified influential members based on a dynamic network and proposed a temporal role affiliation frequency model [11]. On the other hand, they have focused on the motivations for role changes, particularly for single-role transformations, mainly through questionnaire surveys to explore them. For example, Zhao Xin et al. found that professional knowledge and reciprocity norms are internal factors for the transformation from knowledge seekers to knowledge contributors [12], while G. Zeng et al. discovered that self-efficacy, user trust, shared vision, and community loyalty are key to the transformation from lurkers to knowledge contributors [13].

2 Literature Review and Theoretical Foundation

2.1 Virtual Community User Role Research

Virtual community user role research primarily revolves around identifying and classifying user roles based on social network structure characteristics and user behavior features. The former treats users as members of an interaction behavior network, where the network topology of interaction nodes, such as degree centrality and closeness centrality, determines user roles, generally dividing them into information generators, information drivers, and information bridges [4]. Some also classify Q&A platform user roles into learning growth type, helpful type, thoughtful type, silent learning type, and preference uncertain type based on in-degree and out-degree [5]. The latter role identification is usually related to user participation behaviors in the community. For example, early on J. Hagel et al. qualitatively divided community user roles into browsers, lurkers, contributors, and purchasers based on participation level [6]. In subsequent studies, more scientific methods such as statistical analysis and clustering were used to identify user roles, including opinion leaders and experts [7]. For example, J. Villodre redefined influencers, disseminators, and ordinary users based on the number of tweets posted from the perspective of social media and emergency management [8].

2.2 Information Ecology Theory

Information ecology introduces concepts from natural ecology to explore issues related to information systems and information self-organization from an ecological perspective. It is a science that studies information laws [14-15], emphasizing the mutual harmony among information ecological factors. In 1999, B. A. Nardi et al. first proposed the concept of information systems in the book *Information Ecologies: Using Technology with Heart*, considering information ecology as "a system composed of people, practices, values, and technology in a specific environment" [16], calling people, practices, values, and technology information ecological factors. However, scholars have different views on information ecological factors, gradually forming three theories: the two-factor theory (information user and information environment) [17], the three-factor theory (information, information user, and information environment) [18], and the four-factor theory (information, information user, information technology, and information environment) [19]. Information, information users, and the information environment are considered the core elements of the information ecosystem, with information technology being separated from the information environment. Considering that all users entering the community have no difference in access to information technology, this paper mainly applies the three-factor theory to construct the influencing factors of community user role formation.

2.3 Lifecycle Theory

In the study of emergency event lifecycles, classic theories include S. Fink's four-stage model of incubation, outbreak, continuation, and recovery [20] and B. T. Burkholder et al.'s three-stage model of pre-event, during-event, and post-event emergency management [21]. Based on these theories, An Lu et al. compared the differences in severity indicators of various emergencies across different life stages by dividing lifecycles and constructed an early warning mechanism [22]. Liu Bing et al. constructed a risk assessment and decision-making model for major public health emergencies based on S. Fink's lifecycle theory [23]. Jiang Jingui et al. divided the lifecycle into formation, climax, fluctuation, and dissipation periods based on themes and emotional trends [24]. In some emergencies, public opinion lifecycles may evolve in single-peak, double-peak, and multi-peak patterns [25]. In this study, based on the characteristics of public health emergencies, the entire event evolution is divided into incubation, outbreak, first decline, fluctuation, second decline, and calm stages, where the fluctuation period refers to the phenomenon where after the event reaches a certain peak, it experiences a period of silence, and with the stimulation of new information, new peaks appear, presenting an alternating wave pattern.

In summary, previous user role research has mainly focused on static analysis of role identification and classification, while role transformation research only focuses on the transformation of a certain type of role, such as from knowledge seekers to knowledge contributors. There is a lack of analysis of key factors in role formation and dynamic analysis and difference analysis of all role transformations from a holistic perspective, with less attention paid to roles on Q&A platforms. However, user roles and behavioral characteristics differ across different social media platforms. Therefore, this paper aims to identify user roles applicable to Q&A platforms, analyze the key factors in role formation based on information ecology theory and lifecycle theory, and explore the dynamic characteristics and differences of different role transformations.

3 Research Methodology

3.1 Data Collection and Preprocessing

In Q&A platforms, answer quality is often measured by the number of likes [26]. In crisis situations, some scholars also use the number of likes to represent the degree of information recognition or adoption [27]. Since the Zhihu platform does not provide "like" and "favorite" indicators for question answers, this paper uses the number of likes to measure answer quality. Research on question quality is relatively scarce. Li Shengli et al. evaluated question quality based on question scores and question score distribution in CSDN and Stack Overflow communities [28]. Similarly, based on Zhihu platform characteristics, question quality can be measured by the number of likes, number of follows, and number of views a question receives. The number of likes represents the degree of recognition for the question, the number of follows represents the degree of representativeness (i.e., the degree to which it represents people with similar question needs), and the number of views represents the attractiveness of the question. In other words, a high-quality question should gain recognition from other community users, represent the needs of most people, and attract more attention.

This paper collected all questions and answers under the "Novel Coronavirus" topic on the Zhihu Q&A platform from December 30, 2019, to May 31, 2020. After deduplication, a total of 466,274 related questions, answers, and articles were obtained, including 15,401 questions, 976 articles, and 449,897 answers, involving 236,653 real-name users and 42,063 anonymous users. Detailed information for these 236,653 users was further collected based on user IDs, resulting in an initial dataset of 702,927 Q&A and user data entries. Irrelevant questions, answers, articles, and data from deactivated and anonymous users were manually removed, ultimately obtaining 407,247 valid question and answer data entries involving 23,6304 users.

3.2 Q&A Platform User Role Classification

In Q&A platforms, user questioning and answering behaviors are considered the most valuable parts. Drawing on J. Hagel et al.'s idea of qualitatively dividing community user roles from the perspectives of participation level and value [6], this paper reinterprets and divides Q&A platform user roles into lurking, active, demand, and knowledge types, attempting to quantitatively identify user role categories applicable to Q&A platforms. The lurking type corresponds to the lurker proposed by J. Hagel et al., a role that contributes little information to the community. The active type corresponds to the contributor they proposed, usually very active users. The demand and knowledge types correspond to the purchaser they proposed, and are further refined into demand and knowledge types based on Q&A platform characteristics. Unlike Hagel et al., who set participation level and value dimensions to be only related to user behavior, in this paper, participation level refers to the frequency of user questioning and answering in each lifecycle, while value refers to the quality of user questioning and answering in each lifecycle.

The evaluation of answer quality is calculated as shown in Formula (1), where AnswerCount represents the total number of answers by a user, and like_i represents the number of likes for the user's i-th answer:

$$AnswerQuality = \frac{\sum_{i=1}^{AnswerCount} like_i}{AnswerCount} \quad \text{Formula (1)}$$

The evaluation of question quality is calculated as shown in Formula (2), where QuestionCount represents the total number of questions by a user, like_i represents the number of times the user's i-th question is marked as a good question, follow_i represents the number of follows for the user's i-th question, browse_i represents the number of views for the user's i-th question, and w1, w2, w3 are the weights for these three indicators:

$$QuestionQuality = \frac{\sum_{i=1}^{QuestionCount} (w1 \cdot like_i + w2 \cdot follow_i + w3 \cdot browse_i)}{QuestionCount} \quad \text{Formula (2)}$$

This paper combines the entropy weight method and coefficient of variation method to assign variable weights. The former determines weights based on the degree of variation of indicators reflecting their information content, while the latter determines weights by measuring the degree of change in indicator observations. The combination of the two can make the weighting results more accurate. The combined weight used to calculate question quality is shown in Formula (3), where α represents the proportion of the entropy weight method in the combined weight, WE-Wj represents the weight of the j-th indicator calculated by the entropy weight method; 1-α represents the proportion of the coefficient of variation method in the combined weight, WC-Vj represents the weight of the j-th indicator obtained by the coefficient of variation method. Typically, the α coefficient is set to 0.5 [29]:

$$w_j = \alpha \cdot WE - W_j + (1 - \alpha) \cdot WC - V_j \quad j = 1, 2, 3 \quad \text{Formula (3)}$$

Based on user behaviors and quality, for each lifecycle stage, users are identified and classified according to the following principles:

(1) **Lurking Users**: When a user only asks questions, if QuestionQuality ≤ AVG(*{n}^{1} QuestionQuality) and QuestionCount ≤ AVG(* {n}^{1} QuestionCount), meaning both question quality and quantity are below average, the user is a lurking user (n represents the number of users with questioning behavior). Or when a user only answers, if AnswerQuality ≤ AVG(*{m}^{1} AnswerQuality) and AnswerCount ≤ AVG(* {m}^{1} AnswerCount), meaning both answer quality and quantity are below average, the user is also a lurking user (m represents the number of users with answering behavior). This paper defines lurking users as groups that do not actively participate in the community and have low post value based on posting frequency and quality. During crisis response, these users participate in the community but cannot make significant contributions.

(2) **Active Users**: When a user only asks questions, if QuestionQuality ≤ AVG(*{n}^{1} QuestionQuality) and QuestionCount > AVG(* {n}^{1} QuestionCount), meaning question quality is below average but frequency is above average, the user is an active user. Or when a user only answers, if AnswerQuality ≤ AVG(*{m}^{1} AnswerQuality) and AnswerCount > AVG(* {m}^{1} AnswerCount), meaning answer quality is below average but frequency is above average, the user is also an active user (n and m are defined as above). Active users have high activity levels but low output value, often having high social needs [30].

(3) **Demand Users and Knowledge Users**: Demand users have high information needs (high question quality). Knowledge users produce valuable information (high answer quality). The commonality between demand and knowledge users in the community is their high Q&A quality. However, in community platforms, some users both ask and answer questions. To distinguish whether such users lean toward demand or knowledge type, after obtaining all users' question counts, answer counts, question quality, and answer quality for each lifecycle stage, the maximum-minimum standardization method is used to standardize the data to eliminate the impact of indicator value sizes on role division. Drawing on the description of information volume in information theory [31], where the probability of an event is positively correlated with information volume, the information volume calculation formula for event i is H = -$\log_2$P_i. By analogy,

the calculation formula for distinguishing demand and knowledge users is shown in (4). Note that the product of standardized quantity and quality no longer has original meaning because the dimensional impact has been eliminated.

$$R_{demand} = -\log_2(QuestionCount^* \cdot QuestionQuality^* + 1)$$
$$R_{knowledge} = -\log_2(AnswerCount^* \cdot AnswerQuality^* + 1)$$
$$userRole = \begin{cases} \text{Knowledge User} & \text{if } R_{knowledge} \geq R_{demand} \\ \text{Demand User} & \text{if } R_{knowledge} < R_{demand} \end{cases} \quad \text{Formula (4)}$$

3.3 Identification of Influencing Factors in Community User Role Formation

Based on information ecology theory, this paper analyzes the influencing factors of community user role formation from three aspects: information user factor, information factor, and information environment factor, as shown in Table 1 :

Table 1 Influencing Factors from the Information Ecology Perspective

Factor Category	Attribute	Feature	Information Ecology Factor Representation
Information User Factor	Natural Attributes	Gender	Male/Female/Not filled
		Identity	Individual/Organization
		Region	Beijing/Shanghai/Guangzhou…
		Industry	E-commerce/Public Services/Internet…
	Characteristic Attributes	Influence	Ratio of followers to (followers + following + 1)
		Creation Level	[0-10]
		Absolute Role	Questioner/Answerer/Both
		Previous Stage User Role	Lurking/Active/Knowledge/Demand
		Interest Level	Similarity between user topic distribution and environment topic distribution

Factor Category	Attribute	Feature	Information Ecology Factor Representation
Information Factor	Text Attributes	Average Question Title Length	Ratio of total question title length to question count
		Average Question Description Length	Ratio of total question description length to question count
		Average Answer Text Length	Ratio of total answer text length to answer count
		Text Information Volume	Sum of TF-IDF for all words in Q&A text
	Emotional Attributes	Question Title Sentiment Tendency	Ratio of sum of positive sentiment values in question titles to question count
		Question Description Sentiment Tendency	Ratio of sum of positive sentiment values in question descriptions to question count
		Answer Text Sentiment Tendency	Ratio of sum of positive sentiment values in answer texts to answer count
	Temporal Attributes	Average Q&A Time Interval	Ratio of time intervals between all answers and questions to total question and answer counts

Factor Category	Attribute	Feature	Information Ecology Factor Representation
		Time Distribution	Divided into: Late night (00:00-06:00], Early morning (6:00-8:30], Morning (8:30-12:00], Noon (12:00-14:00], Afternoon (14:00-18:00], Evening (18:00-24:00], represented by presence/absence of text posting in each period
	Thematic Attributes	Topic Distribution	Whether a user's text topics include certain topics among all users' text information
Information Environment Factor	Environmental Attributes	Platform Intervention	Ratio of folded answers to (total answers + 1)
		Topic Richness	Total number of topics contained in a user's texts
		Information Discussion Level	Number of question answers, number of answer comments
	Evolution Stage		Incubation, Outbreak, First Decline, Fluctuation, Second Decline, Calm

3.3.1 Information User Factor Information user factors are typically characterized by natural attributes such as user age and gender on social media [32]. This paper also proposes that information user factors include characteristic attributes, which are non-natural attributes but related to the user themselves. Based on Zhihu platform characteristics, the natural attributes of information users are shown in Table 1, where region is re-divided into 36 feature values including 34 Chinese provincial-level administrative regions, overseas, and other regions; industry uses the industry information provided by Zhihu, involving 112 feature values; creation level ranges from 0 to 10, totaling 11 feature values.

The characteristic attributes of information users are shown in Table 1. The absolute role refers to a user's questioning and answering behaviors in a certain stage: if the user only asks questions, they are a questioner; if only answers, an answerer; if both, both questioner and answerer. The previous stage user role is the user's role in the previous life stage. Since users do not always exist in the community, two virtual roles are set: user entry and user exit. If a user enters the community for the first time in a certain life cycle, their previous role is user entry; if the user is not entering for the first time but did not post anything in the previous life cycle, their previous role is user exit. The user's interest level is calculated as shown in Formula (5), representing the similarity between the probability distribution of all text topics posted by a user and the probability distribution of text topics in that life cycle environment. The topic distribution probability is represented by the average of the sum of probabilities of all topics for all users n. This paper uses JS divergence to evaluate their similarity, with a value range of [0,1]. The smaller the value, the more interested the user is.

$$InterestDegree = JS\_divergence(P\{t_1, t_2, \ldots, t_r\}, P) \quad \text{Formula (5)}$$

3.3.2 Information Factor Information factors can be refined into characteristics such as information timeliness and usefulness [33]. Considering that information factors describe the basic characteristics of information, this paper divides information factors into text attributes, emotional attributes, and temporal attributes. Text attributes include not only statistics on the average length of question titles, question descriptions, and answer texts, but also use TF-IDF to calculate the overall information volume of all posts by a user, i.e., using the sum of TF-IDF values of words in all Q&A posts published by a user after removing stop words to represent the information volume of all posts published by the user in a certain life cycle.

For calculating user emotional attributes, this paper uses the Baidu sentiment analysis model Senta with Bi-LSTM to predict post sentiment [34]. First, posts published by users are segmented into sentences, and the positive sentiment tendency probability is used as the text's sentiment tendency value. The calculation is shown in Formula (6), representing the sentiment tendency value of all posts published by a user, where m represents the total number of question or answer posts by the user, n represents the number of sentences in a question or answer post, and P(Sentence_{ji}) represents the positive sentiment probability of the i-th sentence in the user's j-th post.

$$Emotion_{user} = \frac{\sum_{j=1}^{n} \sum_{i=1}^{m} P(Sentence_{ji})}{m} \quad \text{Formula (6)}$$

Thematic attributes include topic distribution and topic richness. This paper uses the Bert-based text clustering tool Bertopic to identify topics in user-posted information. Bertopic uses transformers and c-TF-IDF to create dense clusters, resulting in easily interpretable topics [35]. Topic richness is explained in section 3.3.3.

Temporal attributes include time distribution and dynamic time distribution intervals. The division of time distribution is shown in Table 1, where each user's posts in a life cycle are discretely distributed across the 6 time periods in the table, represented by presence/absence. Regarding dynamic time distribution intervals, user questioning behavior is considered the initial post, so the time interval is 0. User answering behavior time interval is the time difference between the answer and the question it answers. The final dynamic time interval is the ratio of the sum of time intervals between all answers and their questions to the sum of all answer and question counts.

3.3.3 Information Environment Factor Regarding information environment factors, the external environment in social media considers attributes such as forwarding and commenting [30]. This paper's information environment factors consider Zhihu platform intervention, information discussion level, and evolution stage, with specific measures shown in Table 1.

3.4 Analysis of Influencing Factors in User Role Formation

To explore the key factors in user role formation, one-hot encoding is applied to unordered categorical variables in Table 1 (region and industry), and sequential encoding is applied to user gender, user type, verification status, and whether they are an excellent answerer. The user role in that stage is used as the dependent variable, and a multi-classification model is built to identify the best classifier, which is then used as input for SHAP (SHapley Additive exPlanations) training, and feature importance ranking charts for different roles are drawn. SHAP measures model feature importance based on game theory and is interpretable [36]. In this paper, the constructed multi-classifiers include Linear Regression Classifier (LR), K-Neighbors Classifier (KNeighborsClassifier), Neural Network Classifier (MLPClassifier), Decision Tree Classifier (DecisionTreeClassifier), Random Forest Classifier, LightGBM Classifier (LGBMClassifier), CatBoost Classifier, and XGBoost Classifier.

When building the multi-classification model, considering the imbalance in multi-class samples, this paper uses the oversampling method Borderline-SMOTE algorithm to balance the data on the training set. This algorithm can more accurately learn the boundaries of each class, thereby improving the class distribution of samples [37]. To identify the best classifier, Borderline-SMOTE combined with 10-fold cross-validation is used to evaluate different multi-classifiers, i.e., applying the Borderline-SMOTE algorithm to balance the data on the training set in each evaluation, and using the prediction effect on the test set to evaluate the multi-classifier results.

3.5 User Role Transformation Analysis

To mine the patterns of user role transformation, this paper calculates the support and confidence of role transformation, as well as the degree of change in main indicators for different user role transformations between different life cycle stages, as shown in Formula (7), representing the degree of change of a certain influencing factor i for user role U1 transforming to user role U2 in a transformation stage (Stage A - Stage B). Where U2(i) is the mean of attribute i for user role U2 in Stage B, and U1(i) is the mean of attribute i for user role U1 in Stage A. To avoid division by zero, 1 is added to the denominator.

$$DegreeChange(U1,U2)_i = \frac{U2(i) - U1(i)}{U1(i)+1} \quad \text{Formula (7)}$$

In addition, this paper uses the FP-growth association rule algorithm to conduct association analysis on information user factors, information factors, topics, and topic factors under different role transformation conditions to extract potential relationships and discover similarities and differences in different role transformations. FP-growth adopts a divide-and-conquer strategy to accelerate the association rule mining process [38], overcoming the efficiency problems of the Apriori algorithm.

Before conducting association rule analysis, data needs to be categorized. Features such as follower count, following count, influence, average question title length, average question description length, average answer text length, text information volume, and dynamic time distribution interval are divided into two categories based on being above or below the mean. Interest level is divided into four categories using the quartile method, i.e., by values [0,0.25], (0.25,0.5], (0.5,0.75], (0.75,1] into not interested, slightly interested, moderately interested, and very interested. Using Senta's official classification standard, sentiment tendency values are divided into [0,0.45], (0.45,0.55], [0.55,1] into negative, neutral, and positive. Then, by adjusting minimum support and minimum confidence and based on support, confidence, and lift, the most relevant and effective association rules under each role transformation condition are found. The rule with maximum support under maximum lift is selected as the final association rule. The larger the lift, the more relevant the two items are. To simplify the result format, this paper designs some representation rules. Taking "low influence, [C(individual | not excellent answerer)]" as an example, [] indicates that the content inside is optional, and C(A|B) represents all subsets of {A,B} except the empty set. Therefore, this example can form four rules: "low influence," "low influence, individual," "low influence, not excellent answerer," and "low influence, individual, not excellent answerer." If a rule contains a minus sign, it indicates that it does not include any subset of the dataset after the minus sign. For association rules where the antecedent and consequent can be swapped with consistent support, confidence, and lift, they are directly merged and written as "A|B" seed rules.

4 Experiments and Results Analysis

4.1 Data Lifecycle Division

Based on the propagation characteristics of information in cyberspace, the lifecycle is divided by identifying inflection points in the change of information quantity on social media. Considering that posts by anonymous and deactivated users under this topic are also closely related to the topic, only irrelevant data is removed when dividing the lifecycle. The final lifecycle stages are shown in Table 2 , totaling 6 stages.

Table 2 Lifecycle Division

Stage	Date Range	Post Count	Key Event
Incubation	2019.12.30-2020.1.20	1,540	On the evening of Jan 20, Dr. Zhong Nanshan confirmed human-to-human transmission, entering outbreak stage
Outbreak	2020.1.21-2020.2.1	15,401	On Feb 1, Wuhan Huoshenshan Hospital was about to be completed and officially delivered on Feb 2, entering first decline stage
First Decline	2020.2.2-2020.3.11	97,600	On Mar 11, WHO declared Covid-19 a global pandemic, entering fluctuation stage
Fluctuation	2020.3.12-2020.4.4	44,989	On Apr 4, national mourning; global daily new confirmed cases exceeded 100,000, entering second decline stage

Stage	Date Range	Post Count	Key Event
Second Decline	2020.4.5-2020.5.3	407,247	On May 3, National Health Commission: existing confirmed cases nationwide decreased for 11 consecutive days; new asymptomatic cases were 12, the lowest since reporting began, entering calm stage
Calm	2020.5.3-2020.5.31	-	-

4.2 Q&A Platform User Role Classification and Transformation

Based on the method for dividing Q&A platform user roles in section 3.2, we first identified users in each lifecycle stage, ultimately obtaining a total of 289,259 users across all lifecycle stages, including 221,349 lurking users (76.52%), 52,925 active users (18.3%), 14,503 knowledge users (5.01%), and 482 demand users (0.17%). We counted the gender, industry, region, creation level, follower count, following count, and influence characteristics of different roles in each lifecycle stage, finding that users mainly gather in economically developed areas such as Beijing, Shanghai, and Guangdong. Active users have an average follower count stable between 2,000-5,000, with creation level and influence gradually increasing with lifecycle development. Lurking users had about 3,000 followers in the early stage of the epidemic, which dropped to below 1,000 in the middle stage, then slightly rebounded later, with relatively stable following counts and influence levels consistently below 0.4 during the middle stage. Demand users are relatively special. In the early and late stages of the epidemic, the Zhihu official platform autonomously asked questions to promote discussion, resulting in high follower and following counts for demand users, but influence levels around 0.5 and creation levels around 6. Knowledge users have the highest follower counts, influence, and creation levels in all epidemic stages, with influence around 0.8 and creation levels above 6.5. Additionally, in terms of gender, region, and industry, knowledge users have relatively complete information, while lurking users have the most missing values in these features. All types of user roles mainly gather in industries such as internet, clinical healthcare, and higher education, with internet industry users always occupying a place in any role, and clinical healthcare industry users being key users concerned about the epidemic in the early stage.

According to the description of user role transformation in section 3.5, the transformation of various roles is shown in Table 3 . From Table 3, we can see that different roles have different transformation probabilities. Users maintaining an active role have a probability of 49.78%, the probability of changing from active to lurking is 43.88%, while the probability of changing from active to demand or knowledge is much lower. Other role transformation probabilities are specifically shown in Table 3 . Typically, users tend to maintain their original role unchanged, but demand users have a higher probability of transforming to lurking or active, lurking and active users have a relatively high probability of mutual conversion, but active users are more likely to transform to lurking. Demand users have a 13.82% probability of transforming to knowledge, but knowledge users have a low probability of transforming to demand. Additionally, if transformation occurs, users also tend to transform to lurking or active roles. Demand users have different probabilities of transforming to various roles.

Table 3 Role Transformation

Transformation	Probability
Lurking→Lurking	0.32830
Lurking→Active	0.14820
Lurking→Knowledge	0.18901
Lurking→Demand	0.02329
Active→Lurking	0.00077
Active→Active	0.16663
Active→Knowledge	0.29522
Active→Demand	0.49778
Knowledge→Lurking	0.06135
Knowledge→Active	0.43883
Knowledge→Knowledge	0.00204
Knowledge→Demand	0.00118
Demand→Lurking	0.04962
Demand→Active	0.65398
Demand→Knowledge	0.02491
Demand→Demand	0.00059

According to section 3.5, this paper mined the characteristic changes of user role transformations in different lifecycle stages. As shown in Table 4 , due to space limitations, only the top two role transformation probabilities and the top 5 indicators with the highest positive/negative change degrees are displayed. Transformation probability represents the proportion of a certain role transformation count to all role transformation counts in that transformation stage. During different lifecycle transformation periods, the main transforming roles vary slightly, but lurking users maintaining their role unchanged have the highest probability. In each transformation period, the main changing indicators and change degrees for role maintenance and change are also different. Taking lurking role maintenance as an example, in the first two transformation periods, maintaining the role unchanged requires the follower count change degree to be 0.68 times the previous stage, while in later periods such as from second decline to calm stage, maintaining the lurking role unchanged requires the follower count change degree to be 1.95 times the previous stage. From incubation to outbreak stage, maintaining the knowledge role unchanged requires the information volume change degree of posted content to be 0.62 times the previous stage, with posting times in the morning or late night, and more focused topics. In different transformation stages, the requirements for role transformation also differ. Taking the transformation from lurking to active as an example, from outbreak to first decline stage, transforming from lurking to active means the follower count change degree is 1.79 times the previous stage, the question description average length change degree is 3.75 times the previous stage, and the question title average length change degree is 1.55 times the previous stage, with information volume change degree being 0.76 times the previous stage.

4.3 Training and Evaluation of User Role Classification Models

In measuring influencing factor topic distribution, due to the high training cost of Bertopic, after 5 days of training, 1,584 topics were obtained, with a considerable number of documents related to topics only around 10. To ensure each topic has at least 100 documents, combined with topic similarity matrices, the final topic number was determined to be 180, and existing topics were further merged based on topic similarity.

According to the description in section 3.4, Borderline-SMOTE and 10-fold cross-validation are used to evaluate the models. Model evaluation uses macro-averaged precision, recall, F1-score, and accuracy. From Table 5 , it can be seen that both CatboostClassifier and LGBMClassifier have optimal performance in multi-classification, but CatboostClassifier performs better than LGBMClassifier in precision, F1-score, and accuracy. Therefore, when using the SHAP model to mine influencing factors, CatBoostClassifier is used as its input.

Table 5 User Role Multi-Classifier Evaluation Results

Classifier	Precision	Recall	F1-Score	Accuracy
KNeighborsClassifier	0.5942	0.5410	0.7896	0.8233
MLPClassifier	0.7711	0.8737	0.8813	0.8733
DecisionTreeClassifier	0.3201	0.4626	0.6510	0.8035
RandomForestClassifier	0.8544	0.8818	0.8689	0.8834
LGBMClassifier	0.3230	0.4790	0.6603	0.8076
CatBoostClassifier	0.8031	0.8694	0.8729	0.8706
XGBClassifier	0.5508	0.6400	0.8746	0.9147

4.4 Analysis of Influencing Factors in Community User Role Formation

To explore the specific impact of different variables on the formation of different user roles, feature importance ranking charts are drawn (see Figure 1 [Figure 1: see original paper]). In Figure 1, color depth represents feature value size—the lighter the gray, the smaller the feature value; otherwise, the larger.

Figure 1 Feature Importance Ranking for Different User Types

From Figure 1, the following conclusions can be drawn:

(1) **For Lurking Users**: The lower the text information volume, the longer the answer text average length, the shorter the question title average length, the lower the question title sentiment tendency value, the higher the dynamic distribution time interval, the higher the answer sentiment tendency value, the previous stage user role being user entry or lurking, the lower the follower count, uncertain industry and region, the lower the answer comment count, the lower the question answer count, the lower the topic richness, and the later the evolution stage, the more likely the user is to be a lurking user.

(2) **For Active Users**: The higher the text information volume, the lower the answer text average length, the higher the question title sentiment tendency value, the higher the dynamic distribution time interval, the lower the answer sentiment tendency value, the previous stage user role not being user entry, the lower the follower count, uncertain industry and region, the lower the answer comment count, the lower the question answer count, and the lower the topic richness, the more likely the user is to be an active user. Additionally, active users post at any time period except early morning, while lurking users do not like to post in the afternoon, morning, or evening.

(3) **For Knowledge Users**: Information environment factors are an important factor promoting their formation. When answer comment count is higher, answer average text length is longer, question answer count is lower, follower count is higher, information volume is higher, question title average text length is lower, dynamic time distribution interval is lower, topic richness is higher, influence is higher, question title sentiment tendency value is lower, gender tends to be male, and creation level is higher, it promotes users to become knowledge users.

(4) **For Demand Users**: When question answer count is higher, information volume is higher, answer average text length is shorter, question title average length is longer, question title sentiment tendency value is higher, dynamic time distribution interval is shorter, answer comment count is higher, previous stage user role not being user entry, gender tends to be not filled, topic richness is higher, creation level is lower, question description average length is higher, and absolute role attribute not being

answerer, it promotes users to become demand users.

4.5 User Role Transformation Based on Information User and Information Factors

According to section 3.5, this paper constructed association rules for different role transformation conditions as shown in Table 6 . Taking the transformation from lurking to demand as an example, when a user changes from lurking to demand, the user is usually both a questioner and answerer and is very interested in the current topic, with a 78% probability of asking highly informative questions with fewer question title words in the morning. This rule reveals the information user characteristics and information behavior when user roles transform from lurking to demand.

Table 6 Association Rules of Information User and Information Factors for Role Transformation

Association Rule (Information User Factor => Information Factor)	Support	Confidence	Lift
Lurking→Lurking: Low influence, follower count below mean, individual, gender_{not} filled => Answer average text length below mean	0.22	0.80	1.13
Lurking→Active: Low influence, answerer, other region, [C(individual	follower count below mean	not excellent answerer)] => Answer average text length below mean	0.23
Lurking→Knowledge: Follower count above mean => High information volume, answer average text length above mean	0.24	0.70	1.21

Association Rule (Information User Factor => Information Factor)	Support	Confidence	Lift
Lurking→Demand: Both questioner and answerer, very interested, [C(individual	not excellent answerer)] => Dynamic time distribution interval below mean, question title average length below mean, morning, [high information volume]	0.27	0.78
Active→Lurking: Low influence, other region, [C(individual	follower count below mean	not excellent answerer	answerer)] => Answer_{negative}
Active→Active: Follower count below mean, high creation level, answerer, other region, [C(not excellent answerer	individual)] => Answer average text length below mean	0.30	0.82

Association Rule (Information User Factor => Information Factor)	Support	Confidence	Lift
Active→Knowledge: High creation level, high influence, other region, not verified, [individual] => Question title average length above mean, morning, afternoon, [high information volume]	0.38	0.76	1.86
Active→Demand: Both questioner and answerer, follower count below mean, [C(individual	not excellent answerer)] => Answer_{negative}	0.26	0.76
Knowledge→Lurking: Follower count below mean, individual, other industry, [C(not excellent answerer	answerer)] => High information volume, answer_{negative}	0.34	0.71
Knowledge→Active: High influence, answerer, following count below mean, individual, not verified, [not excellent answerer] => High information volume, answer_{negative}	0.32	0.80	1.12
Knowledge→Knowledge: High creation level, follower count above mean, not verified, male, not excellent answerer => Answer_{negative}, answer average text length above mean, [C(evening	high information volume)]	0.33	0.82
Knowledge→Demand: Both questioner and answerer, male, [C(individual	high influence	not verified)] => High information volume, late night	0.33

Association Rule (Information User Factor => Information Factor)	Support	Confidence	Lift
Demand→Lurking: High creation level, other region, not verified, [C(individual	answerer	not excellent answerer)] => Low information volume, dynamic time distribution interval above mean, [answer average text length below mean]	0.20
Demand→Active: Follower count below mean, answerer, other region, [C(individual	not excellent answerer)] => Dynamic time distribution interval above mean	0.26	0.77

Association Rule (Information User Factor => Information Factor)	Support	Confidence	Lift
Demand→Knowledge: Following count below mean, answerer, [C(not excellent answerer	high influence)] => Dynamic time distribution interval above mean, high information volume	0.35	0.75
Demand→Demand: Following count above mean, high creation level, [C(individual	high influence	not verified	not excellent answerer)] => Evening, question title_{positive}, [C(question title average length above mean

Based on Table 6 , the following role transformation patterns are summarized:

(1) **When users transform from different roles to lurking role**: They typically exhibit low answer text length or low information content in their texts, but there are differences in transformations from different roles. When lurking and active users transform to lurking, they often have low

influence—the former tend to have low answer text length, while the latter tend to show low information volume. When knowledge users transform to lurking, their personal follower counts are below mean. This transformation may occur because some opportunity made the user a knowledge user in a certain stage but they cannot sustain it. When demand users transform to lurking, they often have high creation levels and show dynamic time distribution intervals above mean and low information volume in behavior, indicating that although they are demand users, they will answer some questions after absorbing some knowledge, but their answering behavior often has lag and the information volume is not high.

(2) **When users transform from different roles to active role**: Lurking and active users transforming to active show the same behavior pattern—low answer text length—but the former often has low influence while the latter often has high creation level. Knowledge and demand users show completely different information user characteristics and behavior patterns when transforming. When knowledge users transform to active, they are usually high-influence users answering texts with high information volume and relatively negative sentiment, while when demand users transform to active, they have fewer followers and directly show high answer lag in their answers.

(3) **When users transform from different roles to knowledge role**: High information volume in text is a universal principle, but active users transforming to knowledge is an exception. When active users transform to knowledge, they have high influence and high creation level, with an 85% probability of answering relatively negatively, and the information volume of answer text is less important to them. When lurking users transform to knowledge, their follower counts are above mean, with a 70% probability of answering longer texts with high information volume. When users maintain the knowledge role unchanged, they often have high creation level, follower counts above mean, and high influence, with an 80% probability of answering content with high information volume and negative sentiment. When demand users transform to knowledge, if their following count is below mean, their answers, although lagging, contain high information volume; if following count is above mean, they prefer to post information content in the afternoon or evening, which also often has high information volume.

(4) **When users transform from different roles to demand role**: They show different posting time tendencies and text length tendencies, as well as the same high information volume text characteristic. When lurking, active, and knowledge users transform to demand, they usually share the common characteristic of being both questioners and answerers. When lurking users transform to demand, when they are very interested in the topic, they will ask or answer in the morning, with question title length below mean and answer time interval below mean. When active users

transform to demand, follower counts are often below mean, with a 76% probability of asking or answering in the morning and afternoon, and question title length above mean. When knowledge users transform to demand, they are high-influence users who tend to ask or answer questions late at night or in the evening. When users maintain the demand role unchanged, they have high following counts, high creation level, and high influence, with an 80% probability of asking questions in the evening, relatively positive question titles, and question title average length above mean.

4.6 User Role Transformation Analysis Based on Themes

As shown in Table 7 , the thematic association rules that users focus on under each role transformation condition are presented. Due to space limitations, only thematic association rules with lift greater than 2.5 are displayed. Based on thematic user role transformation analysis, taking the transformation from active users to demand users as an example, it can be understood that when users transform from active to demand, they are most likely to simultaneously focus on two topics: herd immunity and the continuous rise in U.S. death toll. Users focusing on herd immunity have a 100% probability of also being interested in the continuous rise in U.S. death toll. Under all role transformation conditions, the topics users focus on basically revolve around political wars triggered by the U.S., collapse of the U.S. medical system, herd immunity, Bill Gates' personal situation, continuous rise in U.S. death toll, U.S. infections, confirmed cases and death toll, etc. Under different role transformation conditions, the topics users are most likely to focus on also differ. For example, under the condition of lurking users transforming to demand users, 38% of users simultaneously focus on the collapse of the U.S. medical system, Spanish flu and U.S. flu outbreak, [C(herd immunity | continuous rise in U.S. death toll)]. If users mainly focus on "collapse of the U.S. medical system," under this transformation condition, users have a 100% probability of being interested in Spanish flu and U.S. flu outbreak, [C(herd immunity | continuous rise in U.S. death toll)]. Under the condition of demand to demand transformation, users will simultaneously focus on coronavirus antibody research and Covid-19 vaccine entering clinical trials, or focus on "discrete Jiangsu" supporting various places and praying for people to be safe and get through difficulties to see sunshine, etc. Other rules are similar. By analyzing the topics users focus on under different role transformation conditions, we can understand the content of concern during different user transformations in public health emergencies, which has certain significance for platforms to timely push related Q&A. For example, when users maintain the lurking role unchanged, we can observe the content of concern when lurking users transform to active, and by calculating the similarity between lurking and lurking users, push similar Q&A to stimulate lurking users to transform to active.

Table 7 Thematic Association Rules for Role Transformation (Partial)

Association Rule (Topic => Topic)	Support	Confidence	Lift
Lurking→Active: Continuous rise in U.S. death toll, U.S.-triggered political war, U.S. economy => Sarcastic cheering for U.S., Bill Gates personal situation, New Zealand	0.33	0.85	1.07
Lurking→Demand: Collapse of U.S. medical system => Spanish flu and U.S. flu outbreak, [C(herd immunity	continuous rise in U.S. death toll)]	0.38	0.76
Active→Active: U.S.-triggered political war, New Zealand prevention strategy => Sarcastic cheering for U.S., Bill Gates personal situation	0.26	0.76	1.18
Active→Demand: Herd immunity => Continuous rise in U.S. death toll	0.34	0.71	1.08
Knowledge→Knowledge: Continuous rise in U.S. death toll, U.S. economic shutdown => Sarcastic cheering for U.S., Bill Gates personal situation, U.S.-triggered political war	0.32	0.80	1.12
Knowledge→Demand: Both questioner and answerer, male, [C(individual	high influence	not verified)] => High information volume, late night	0.33
Demand→Lurking: High creation level, other region, not verified, [C(individual	answerer	not excellent answerer)] => Low information volume, dynamic time distribution interval above mean, [answer average text length below mean]	0.20

Association Rule (Topic => Topic)	Support	Confidence	Lift
Demand→Active: Follower count below mean, answerer, other region, [C(individual	not excellent answerer)] => Dynamic time distribution interval above mean	0.26	0.77
Demand→Knowledge: Following count below mean, answerer, [C(not excellent answerer	high influence)] => Dynamic time distribution interval above mean, high information volume	0.35	0.75
Demand→Demand: Following count above mean, high creation level, [C(individual	high influence	not verified	not excellent answerer)] => Evening, question title_{positive}, [C(question title average length above mean

5 Conclusion and Future Work

This paper takes the "Novel Coronavirus" topic on the Zhihu platform as an example. First, from the dimensions of participation level and value, user roles in different lifecycle stages of the Q&A platform are identified and divided into lurking users, active users, demand users, and knowledge users. Then, based on information ecology theory, a multi-classifier model (optimal: Catboost) for community user role formation is constructed from information user, information, and information environment factors, and the SHAP model is used to determine the key influencing elements for different user roles. Finally, association rule methods are used to analyze user role transformation behavior patterns and topic characteristics based on information user factors and information factors, revealing user role evolution behaviors and Q&A topics in the context of public health emergencies.

The main conclusions are: (1) Across all transformation stages, users tend to maintain their roles unchanged, with active users having the highest probability of maintaining their role at 49%. Other roles mainly transform to active and lurking users. From different transformation stages, taking the most probable lurking-lurking user (i.e., role maintenance) as an example, the characteristic change degrees in different transformation stages show significant differences. To maintain the role unchanged, the follower count change degree in the later stage of epidemic development is typically required to be 1.95 times the previous stage. (2) The key factors affecting different user roles are inconsistent, but information volume is always an important feature. Lower information volume 反而更能说明该用户是潜水型用户. The temporal attribute of information factors is a significant feature of active users, indicating that active users do not care about time periods when participating in the community. The information environment has a greater impact on the formation of demand and knowledge users, followed by text and emotional attributes in information factors. (3) When users transform to different roles, they have different transformation patterns and performance characteristics. For example, when transforming from active to demand, follower counts are often below mean, with a 76% probability of asking or answering in the morning and afternoon, and question title length above mean. Users are also most likely to simultaneously focus on topics of herd immunity and the continuous rise in U.S. death toll.

The innovations and contributions of this research lie in: (1) Theoretical value: It provides a user segmentation model and research method in the context of infectious disease outbreaks; constructs influencing factors of user role formation from the information ecology perspective, reveals key elements affecting different role formations, and expands the application of information ecology theory in social Q&A platforms; explores user role transformation, behavior patterns, and topic characteristics based on association rules, deepening user role research. (2) Practical implications: For Q&A communities, platforms can provide personalized services based on user role behaviors and topic characteristics to promote benign role transformation. For example, recommending knowledge users with similar topics to lurking users can increase the probability of transforming to active users. Platforms can also improve the proportion of important user types based on main factors in user role formation. The study shows that a good information environment and thematic content—i.e., a lively discussion atmosphere and rich topics—are key to the formation of demand and knowledge users. For governments, the behaviors and topic characteristics of users on community Q&A platforms during Covid-19 revealed in this study can help public opinion management departments understand user transformation mechanisms, especially the behaviors and topic characteristics of different role types transforming to demand users, to avoid public opinion risks by timely meeting information needs.

This study has certain limitations: (1) The user role transformation studied covers all lifecycle stages, without specifically subdividing and distinguishing user role transformation characteristics in different lifecycle stages; (2) Limited by data feature acquisition from the Zhihu platform, the indicators used to measure question answer quality are relatively simple, and the existence of information epidemics may cause certain errors in the research; (3) This paper only studies user role identification, formation, and transformation under the Covid-19 topic on the Zhihu platform. The applicability and generalizability of research conclusions to other event contexts need further exploration. Future research will continue to study user role formation and transformation patterns on more Q&A platforms in other infectious disease contexts. Additionally, more association rules can be mined based on the influencing factors in this paper, but this paper only discusses behaviors and topics without analyzing other associated factors. Therefore, future research can further explore these issues.

Acknowledgments: Thanks to the National Experimental Teaching Demonstration Center of Library and Information Science for providing experimental support for this research!

References

[1] People's Daily. Xi Jinping's speech at the meeting of the Standing Committee of the Political Bureau on responding to the novel coronavirus pneumonia epidemic [EB/OL]. [2022-02-15]. http://cpc.people.com.cn/n1/2020/0215/c64094-31588554.htm

[2] Fang Chencheng, Zhang Jiantong. Review and prospect of user research in social Q&A communities [J]. Journal of Intelligence, 2018, 37(9): 185-193.

[3] Zhang Weiwei, Zhu Jie, Jiang Xue. Research on the influence of social learning on knowledge contribution behavior of different types of users in professional virtual communities [J]. Information and Documentation Services, 2021, 42(5): 94-103.

[4] ZHOU X, WU B, JIN Q. User role identification based on social behavior and networking analysis for information dissemination [J]. Future generation computer systems, 2019, 96: 639-648.

[5] Chen Ye, Wang Le, Chen Tianyu, et al. Research on user portraits of social Q&A platforms based on social network analysis [J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(4): 414-423.

[6] HAGEL J, ARMSTRONG A. Net gain: expanding markets through virtual communities [M]. Boston: Harvard Business Press, 1997.

[7] Zhang Shusen, Liang Xun, Qi Jinshan. Survey of social network role identification methods [J]. Chinese Journal of Computers, 2017, 40(3): 649-673.

[8] VILLODRE J, CRIADO J I. User roles for emergency management in social media: understanding actors' behavior during the 2018 Majorca Island flash floods [J]. Government information quarterly, 2020, 37(4): 101521.

[9] PREECE J, SHNEIDERMAN B. The reader-to-leader framework: motivating technology-mediated social participation [J]. AIS transactions on human-computer interaction, 2009, 1(1): 13-32.

[10] FU C. Tracking user-role evolution via topic modeling in community question answering [J]. Information processing & management, 2019, 56(6): 102075.

[11] BARTAL A, RAVID G. Member behavior in dynamic online communities: role affiliation frequency model [J]. IEEE transactions on knowledge and data engineering, 2019, 32(9): 1773-1784.

[12] Zhao Xin, Wang Qianwen, Zhang Changzheng. Research on the mechanism of user role transformation from knowledge seekers to knowledge contributors in professional virtual communities [J]. Information Science, 2017, 35(10): 18-22.

[13] ZENG G, GUAN H, CHEN F. Knowledge sharing in a virtual community of a hotel association: from free riders to active knowledge sharers [J]. Journal of China tourism research, 2014, 10(1): 16-16.

[14] ERYOMINA L. Information ecology-viewpoint [J]. International journal of environmental studies, 1998, 54(3/4): 241-253.

[15] Ma Haiqun, Li Zhongjun, Zhang Tao. Formation, succession and operation of data news information ecology chain [J]. Library and Information Work, 2021, 65(5): 118-125.

[16] NARDI B A, O'DAY V. Information ecologies: using technology with heart [M]. Cambridge: MIT Press, 1999.

[17] Lou Cequn, Zhao Guiqin. The role of information ecological balance in building a harmonious society [J]. Information Science, 2006(11): 1606-1610.

[18] Zhang Haitao, Weng Yuqi, Liu Kuo, et al. Ecological information: connotation, characteristics, movement process and operation law [J]. Library and Information Work, 2013, 57(12): 46-50.

[19] Dou Yue. Research on the construction of "3$\times$3" emergency intelligence system from the perspective of information ecology [J]. Library and Information Work, 2020, 64(15): 82-89.

[20] FINK S. Crisis management: planning for the inevitable [M]. New York: American Management Association, 1986: 20.

[21] BURKHOLDER B T, TOOLE M J. Evolution of complex disasters [J]. The lancet, 1995, 346(8981): 1012-1015.

[22] An Lu, Chen Miaomiao, Li Gang. Research on severity assessment and early warning mechanism of emergencies in social media environment [J]. Library and Information Work, 2021, 65(5): 98-109.

[23] Liu Bing, Xiao Gaofei, Chao Shiyu. Research on risk assessment and decision-making model for major public health emergencies [J]. Journal of Information Resources Management, 2021, 11(5): 17-26, 37.

[24] Jiang Jingui, Yan Siqi. Research on microblog public opinion evolution based on the interaction of theme and emotion: taking the "Red Yellow Blue child abuse incident" as an example [J]. Journal of Intelligence, 2018, 37(12): 118-123.

[25] An Lu, Wu Lin. Analysis of microblog public opinion evolution integrating theme and emotion features [J]. Library and Information Work, 2017, 61(15): 120-129.

[26] Chen Juan, Gao Shan, Deng Shengli. Analysis of user characteristics identification and behavior motivation in social Q&A: taking Zhihu as an example [J]. Information Science, 2017, 35(5): 69-74, 80.

[27] Han Shixi, Zeng Yueliang. Research on influencing factors of digital youth's WeChat public platform health information adoption intention in the context of public health emergencies [J]. Library Science Research, 2021(6): 83-92.

[28] Li Shengli, Zhong Ying. Empirical comparative study of Chinese and foreign technical Q&A communities and its enlightenment: taking CSDN and Stack Overflow as examples [J]. Journal of the China Society for Scientific and Technical Information, 2020, 39(9): 989-1000.

[29] Han Meng, Wu Hong, Li Chang, et al. Research on identification of transferable patents in universities: based on Bayesian theory and combination weighting method [J]. Library and Information Work, 2021, 65(5): 118-125.

[30] RIDINGS C, GEFEN D, ARINZE B. Psychological barriers: lurker and poster motivation and behavior in online communities [J]. Communications of the Association for Information Systems, 2006, (1): 95-119.

[31] SHANNON C E. A mathematical theory of communication [J]. Bell system technical journal, 1948, 27: 370-423.

[32] Zhao Dan, Wang Xiwei, Li Shimeng, et al. Research on network public opinion characteristic quantity and behavior law in new media environment: based on information ecology theory [J]. Journal of the China Society for Scientific and Technical Information, 2017, 36(12): 1224-1232.

[33] Li Jiaxing, Wang Xiwei, Li Shimeng, et al. Research on influencing factors of WeChat usage behavior of elderly user groups from the perspective of information ecology [J]. Library and Information Work, 2017, 61(15): 25-33.

[34] TIAN H, GAO C, XIAO X, et al. SKERP: sentiment knowledge enhanced pre-training for sentiment analysis [J]. ArXiv preprint arXiv:2005.05635, 2020.

[35] GitHub. BERTopic [EB/OL]. [2021-12-04]. https://github.com/MaartenGr/BERTopic

[36] GitHub. SHAP [EB/OL]. [2022-03-15]. https://github.com/slundberg/shap

[37] HAN H, WANG W Y, MAO B H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning [C]//International conference on intelligent computing. Berlin: Springer, 2005: 878-887.

[38] BORGELT C. An implementation of the FP-growth algorithm [C]//Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations. New York: Association for Computing Machinery, 2005: 1-5.

Author Contributions: Chen Miaomiao: proposed research ideas, responsible for literature collection, research design, data collection and processing, writing and revising the paper; An Lu: guided research ideas and methods, proposed revision suggestions, revised the paper.

Publication Ethics Statement

To strengthen and enhance academic norms, research integrity, and academic ethics in the entire process of academic paper writing, reviewing, and editing, establish good academic atmosphere, promote scientific spirit, resolutely resist academic misconduct, and establish and maintain a fair, just, and open academic exchange ecological environment, the *Library and Information Work* Magazine (including the editorial departments of *Library and Information Work* and *Knowledge Management Forum*) has formulated this publication ethics statement based on the actual situation of the two journals and officially released it in February 2020.

This publication ethics statement commits that the two journals will strictly abide by and implement national policies and regulations related to academic ethics and editing and publishing, standardize the behaviors of authors, peer reviewers, and editors in the entire editing and publishing process, and accept supervision from the academic community and the whole society. It includes three major parts with a total of fifteen articles: 1. Author's publication ethics (Academic papers are an important part of scientific research; Academic misconduct is a cancer of academic papers; Authors are the main contributors to academic papers; Author 署名 reflects intellectual property rights and academic contributions; Academic papers should attach great importance to intellectual property rights and information security; Normative citation of references is an important representation of academic norms; High attention should be paid to the standardization of research data and management; Establish error correction and academic self-purification mechanisms). 2. Peer review expert's publication ethics (Peer review is an important quality control mechanism for papers; Review experts should comply with relevant requirements for paper review; Review experts should strictly follow relevant ethical guidelines and codes of conduct). 3. Editor's publication ethics (Editors should be guardians of academic quality; Editors should play a monitoring role in academic ethics construction; Editors should become the last barrier to curb academic misconduct; Zero tolerance for academic misconduct).

Full text available at: http://www.lis.ac.cn/CN/column/column291.shtml

Library and Information Work Magazine Publishing Ethics Statement (Journal News)

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