Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Uncovering How Information Quality Shapes Diverse User Engagement on Content Community Platforms: Harnessing Deep Learning for Feature Extraction

Yanzi Wang [1,2], Min Wang [3,*], Zhen Zhu [1,*] and Haoxiang Ye [1]

1 School of Economics and Management, China University of Geosciences, Wuhan 430078, China; agathaice@cug.edu.cn or 3092059@nyist.edu.cn (Y.W.); wahaha5@cug.edu.cn (H.Y.)
2 Fanli Business School, Nanyang Institute of Technology, Nanyang 473004, China
3 School of Business Administration, Nanchang Institute of Technology, Nanchang 330044, China
* Correspondence: 2023994939@nit.edu.cn (M.W.); zhuzhen@cug.edu.cn (Z.Z.)

Abstract: The success of content community platforms (CCPs) heavily depends on the active engagement of users attracted by externally generated content. Previous research has highlighted the differentiation among various forms of user engagement, such as likes, comments, and retweets, in shaping the dynamics of value co-creation on CCPs. Our objective is to uncover distinct patterns of impact that information quality features have on these different forms of user engagement. Specifically, we employed deep learning techniques to extract information quality features and identified them as persuasive factors operating through central and peripheral routes based on the elaboration likelihood model (ELM), stimulating user engagement. Our dataset was derived from MaBeeWoo, China's largest specialized CCP for travelogues with minimal barriers for creating text and image-based travelogues. By utilizing a negative binomial model, our analysis reveals significant differences in antecedents between retweets and likes/comments while also highlighting variations in the impact levels of specific content quality features between likes and comments. These findings suggest contrasting patterns regarding how content quality features influence information production and dissemination on CCPs, underscoring the necessity for platform sponsors to develop adaptive mechanisms aligned with their strategic objectives for incentivizing specific quality features.

Keywords: content community platform; information quality; user engagement; deep learning

Citation: Wang, Y.; Wang, M.; Zhu, Z.; Ye, H. Uncovering How Information Quality Shapes Diverse User Engagement on Content Community Platforms: Harnessing Deep Learning for Feature Extraction. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 2673–2693. https://doi.org/10.3390/jtaer19040128

Academic Editor: Mirjana Pejić-Bach

Received: 1 July 2024
Revised: 25 August 2024
Accepted: 6 September 2024
Published: 4 October 2024

1. Introduction

Content community platforms (CCPs) are digital platforms that facilitate interactive engagement between content creators and consumers, allowing for the sharing of user-generated content in various formats such as text, images, and videos [1,2], commonly without mandating real-name registration requirements. In comparison to microblogging platforms like Twitter and review platforms such as Yelp, CCPs have a more relaxed length restriction limit and foster a culture of detailed and rich content [3]. Consequently, prominent forms of CCPs have emerged as essential channels for individuals to access information alongside the flourishing of content consumption and entrepreneurship; these include knowledge-sharing communities (e.g., Zhihu) [4], media-sharing platforms (e.g., YouTube) [5], and Q&A communities (e.g., Stack Overflow) [6]. The growth of CCPs greatly depends on their capacity to engage users amidst intensifying competition among these platforms [7,8]. The leading Chinese video CCP platform, Bilibili.com, has successfully implemented interactive features such as likes, coins, and bookmarks to foster an engaging culture known as the "one-click trifecta". In contrast to this success story, Weishi, a short video CCP operated by Tencent.com, experienced two shutdowns in 2017 and 2021 due to its inability to consistently stimulate user engagement despite receiving substantial traffic support from Tencent.com. Consequently, CCPs provide financial and reputational incentives to content creators based on their capacity to attract user engagement [6,9,10].

Although promoting increased user engagement is crucial, the sponsor of a CCP still faces the challenge of striking a balance between attracting diverse forms of user engagement (such as likes, comments, and retweets) due to their varied origins and functions in shaping value co-creation on CCPs [11]. Previous studies have demonstrated that "liking" reflects a positive attitude towards content and demonstrates social support for content creators, thereby encouraging their continued contributions [12,13]. By contrast, leaving a comment indicates a more complex attitude towards the content, which can be advantageous in generating novel insights and fostering the creation of new content [13,14]. Conversely, retweeting, as a mechanism for disseminating messages, can amplify the broader reach of the original content while also reducing participation costs and facilitating network expansion within CCPs [15,16]. Therefore, platform sponsors should establish specific objectives for different forms of user engagement while considering the trade-offs among various outcomes and responding to diverse user engagements [13,17]. This necessitates providing effective guidance to content creators in aligning their content with the strategic interests of platform sponsors. However, the existing literature lacks explicit recommendations for this issue.

The engagement capacity of content is undeniably reliant on its quality [1,18–20]. Nevertheless, content found on CCPs often tends to be extensive and informative—for instance, crowd-testing reports [21] and knowledge-sharing posts [4]. Furthermore, the expertise level, along with motivation and domain knowledge possessed by creators, exerts a substantial impact on the overall quality [22]. In this fiercely competitive environment where an abundance of content exists, creators may prioritize incorporating visually appealing elements rather than focusing solely on intrinsic quality within the content [23,24]. Consequently, sponsors face challenges in managing the overall quality standards within their respective CCPs. Previous research has predominantly adopted a consumer-centric perspective and has focused on investigating the impact of contextual quality (e.g., timeliness, richness, relevancy) [1,18,25] and representational quality (e.g., understandability, format, image) [26–28] of content [24]. These studies heavily relied on consumers' perceptual surveys of content quality and were criticized for overlooking the significant influence exerted by content creators [22]. Given the advancement of content engineering [29], text mining techniques and classification algorithms have been widely employed to extract intrinsic features related to content quality [4,19,21,30–32]. However, existing applied schemes are inadequate for lengthy or informative content on CCPs due to their limitations in effectively addressing semantic complexity.

Jointly, this study aims to provide a comprehensive understanding of the impact of content quality on distinct engagement behaviors (i.e., likes, comments, and retweets) among content consumers on CCPs. This research is conducted within the context of the MaBeeWoo platform, which is the largest specialized CCP for travelogues in China. MaBeeWoo serves as a dynamic community dedicated to sharing travel experiences by nonexpert creators, facilitating the unrestricted creation of travelogues with diverse textual and visual content while allowing users to engage without real-name registration. User engagement plays a crucial role in earning prestigious reputation badges on MaBeeWoo.

This study integrates existing research on consumer engagement behavior (CEB) and information quality (IQ), making significant contributions to the effective management of content quality on CCPs. It reveals contrasting impact patterns of content quality features on information production and dissemination on CCPs, highlighting the necessity for developing adaptive mechanisms by platform sponsors with different strategic objectives in promoting distinct content generation, diversity, or influence to incentivize specific content quality features within the content.

The remaining sections of this paper are structured as follows: Section 2 provides a comprehensive literature review on CEB and IQ, presenting a framework for distinguishing between three forms of user engagement. Subsequently, Section 3 proposes seven hypotheses, followed by an approach for extracting IQ features using deep learning algorithms in Section 4. Section 5 presents empirical results generated from the application of the negative binomial model. Finally, Section 6 concludes with a discussion of the main findings and their implications for theory and practice.

2. Related Work

2.1. User Engagement on CCPs

The term "user engagement" in this paper primarily refers to the behavioral engagement of content consumers. Consumer engagement behavior (CEB) is defined as a behavioral construct that goes beyond purchase behavior, focusing on a specific entity and driven by motivational factors [17,33,34]. The most prevalent CEBs on CCPs are likes, comments, and retweets [11,13,34,35]. While previous studies have often treated them as interchangeable measures of user engagement [13], this paper explicitly distinguishes them as three conceptually distinct forms of engagement, as illustrated in Table 1.

Table 1. Comparison between likes, comments, and retweets.

User Engagement	Motivation	Consumer Input	Description	Emotional State	Cognitive Effort	Value Co-Creation
Like	entertainment	production	self-expression	Positive	low	Information Production
Comment	information	participation	opinion sharing	Complex	high	Information Production
Retweet	social-connection	consumption	content sharing	Complex	low	Information Dissemination

Expanding on the research conducted by Shao [35] and Courtois et al. [36], Heinonen [11] proposed a 3 × 3 matrix to delineate various types of CEBs by considering consumers' input (i.e., consumption, participation, and production) and motivation (i.e., information, social connection, and entertainment). Within Heinonen's matrix, self-expression is associated with the entertainment × production cell, while opinion sharing corresponds to the information × participation cell. These two forms of CEBs indicate liking and commenting, respectively, as suggested by Oh et al. [34]. Heinonen [11] also recognized content-sharing among community members as a correlated CEB for the cell representing social-connection × consumption. This CEB can be likened to a retweet on CCPs, which optimizes message dissemination by amplifying their speed and reach through convenient content resharing [37].

Specifically, in the field of marketing, consumer engagement is considered a multidimensional construct encompassing cognitive, emotional, and behavioral dimensions and plays a pivotal role in relational exchange processes [38]. Yang et al. [13] have provided a comprehensive distinction between likes and comments based on the expressed emotional state and required cognitive effort [38]. They argue that likes require less cognitive effort as a lightweight feedback action for expressing positive emotions. By contrast, comments are a deliberate form of communication that demands more cognitive efforts, allowing for the expression of complex emotions such as admiration, disagreement, or a fusion thereof. Therefore, written comments are consistently regarded as a higher level of CEB than likes [39,40]. Building upon the insights from Yang et al. [13], we posit that the act of retweeting enables a more profound expression of intricate emotions due to retweeters sharing content with online acquaintances [15]. However, since it serves as a channel for message dissemination, content consumers are not required to express their attitudes in writing but can simply click on pre-designed features on the interface, thereby necessitating less cognitive effort compared to commenting.

Additionally, the actions of liking, commenting, and retweeting play distinct roles in value co-creation on CCPs [11,14]. The action of liking indicates users' approval of the content and can be perceived as social support for content creators on CCPs [12], thereby incentivizing them to continue contributing. Commenting allows users to share their opinions with content creators and other participants, generating novel insights that support diverse content creation [14]. Hence, likes and comments can be regarded as mechanisms for information production on CCPs [15]. In comparison, retweeting significantly amplifies the broader reach of the original content [16] while also reducing participation costs and facilitating network expansion within CCPs [15].

2.2. Information Quality of CCPs

Information quality (IQ) is considered essential in achieving semantic success for an information system [41,42]. Wang and Strong's conceptual framework defines four dimensions that encompass IQ: intrinsic, contextual, representational, and accessibility qualities [24]. However, accessibility is often overlooked, as it is typically ensured by CCPs through user-friendly features. Each of these dimensions can be further elaborated upon, as illustrated in the main line of Table 2 [43].

Table 2. Research on IQ of CCPs.

	Dimensions of IQ			Methodology
	Intrinsic	**Contextual**	**Representational**	
[43]	Believability; Accuracy; Objectivity; Reputation	Value-added; Relevancy; Timeliness; Completeness; Amount of data	Interpretability; Ease of understanding; Representational consistency; Concise representation	Meta-Analysis
[26,27]	Accuracy	Currency; Relevancy; Usefulness; Completeness	Ease of Understanding; Format	Field Experiment
[1,18,25]	Reliability; Objectivity; Credibility	Value-added; Timeliness; Richness; Novelty; Completeness	Format; Clear	Questionnaire
[4,19,21,30,31]	Sentiment; Reputation	Relevancy; Information Richness	Readability	Text Mining and Classification Algorithm
[32]	Accuracy; Objectivity; Source Credibility	Information Richness; Relevancy; Timeliness; Quality and quantity of replies	Ease of Understanding	
This Paper	Believability; Objectivity; Content reputation; Source reputation	Relevancy; Richness	Readability	Text Mining and Classification algorithm and Deep Learning

Conventionally, IQ refers to the attributes of information that align with users' expectations [44]. In accordance with this consumer-centric perspective on IQ, previous studies have frequently employed field experiments [26,27] and questionnaires [1,18,25] to extract detailed features of IQ that are readily perceivable by content consumers. However, these approaches particularly emphasize the contextual and representational dimensions of IQ and are limited in capturing intrinsic quality due to its susceptibility to self-reporting bias. Lukyanenko et al. [22] also criticized the consumer-centric perspective of IQ potentially overlooking the significant influence exerted by individual content creators' abilities, motivation, and domain knowledge on CCPs, given the characteristics of crowdsourced content creation on CCPs.

With the growing emphasis on content marketing [29], a content-centric perspective has emerged, leading to the utilization of text mining techniques and classification algorithms for extracting the intrinsic features of IQ. In this context, content sentiment can be extracted as either neutral or affective (positive/negative), which has proven to be significant in influencing user engagement [19,21,30,32]. Moreover, the measurement of information richness and readability can involve the extraction of key concepts and intricate terms from the content [21,30,32]. For instance, Goh et al. [30] computed the information richness of brand community content by extracting key concepts conveyed through advertisements such as product price and quality, subsequently determining the net valence of such content as the disparity between positive and negative sentiments associated with these specific concepts. Notably, Liu et al. [32] developed a text analytic framework specifically tailored for extracting IQ features from CCPs. However, existing approaches have not considered the potential bias introduced by variations in content length. In lengthy content like crowd-testing reports [21] and knowledge posts [4], the key concepts can be significantly more abundant and intricate than what has been extracted so far. This limitation can impede accurately describing content features and drawing robust conclusions about the relationship between IQ and user engagement.

3. Conceptual Framework and Hypotheses

In the investigation of the impact of IQ on user behavior, the elaboration likelihood model (ELM) is commonly employed in the existing literature [26,45]. ELM is a theoretical framework within social psychology that aims to elucidate how individuals internalize external information, thereby leading to the formation or alteration of attitudes [46]. The model posits two distinct routes to persuasion in the process, namely the central route and peripheral route, based on the information receivers' distinct cognitive effort that needs to be expended on information processing [47]. When a receiver is highly motivated and competent, they engage in proactive and careful processing, leading to high elaboration likelihood via the central route. Conversely, when a receiver lacks motivation or competence, their elaboration likelihood decreases and they rely on peripheral cues such as source credibility or heuristics for simple information processing via the peripheral route [48]. Since the ELM does not exclude the possibility of multi-channel information processing [49], attitude formation or change in content consumers on CCPs can occur as a result of both central and peripheral routes [50,51].

Drawing upon the ELM framework and the multi-dimensional structure of IQ, we have developed a research model, illustrated in Figure 1. The role of IQ on CCPs is manifested through the central route, while source credibility acts as a cue influencing user engagement via the peripheral route. Following the text analytic framework proposed by Liu et al. [32] and previous studies, we selected representative features to assess IQ across intrinsic (believability and objectivity) [4], contextual (relevancy and information richness) [21,30,31], and representational (readability) [4,21] dimensions. These incorporated quality features have been recognized by Agarwal and Yiliyasi [52] as posing potential challenges for the IQ of CCPs. To evaluate source credibility, we consider both content and source reputation [21,32]. Henceforth, our focus lies in examining how these IQ features can predict user engagement on CCPs through three main forms: likes, comments, and retweets.

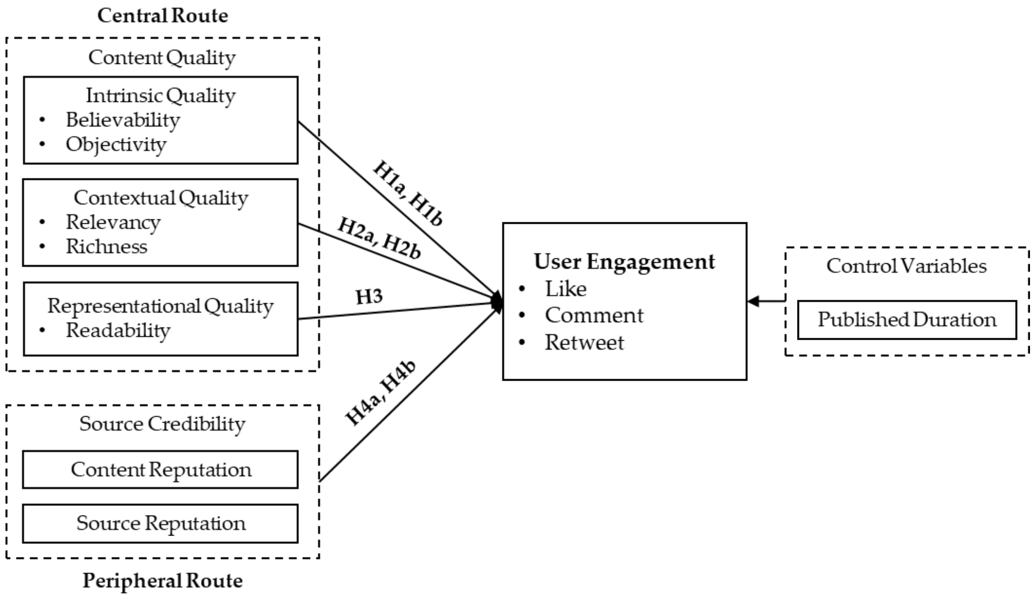

Figure 1. Research model and hypotheses.

3.1. Intrinsic Quality and User Engagement

The intrinsic quality of information refers to its inherent attributes, regardless of its application context [42,43]. Building on the work of Liu et al. [32] and Jin et al. [4], we further defined intrinsically qualified content on CCPs as believable and objective.

Deception on CCPs stands out due to the significant advantage that can be gained from falsehoods and the diminished interpersonal awareness resulting from the isolation and relative anonymity of content creators [53]. Particularly in unstructured textual or visual communication, content creators are constrained when conveying their intentions through a combination of spoken words, tone, and facial expressions due to physical separation from content consumers. Consequently, content consumers encounter challenges in ascertaining creators' inherent goals, mood, or motives [54]. Although distinct user engagement may arise from various antecedents, we posit that believability acts as a determining factor encouraging all forms of user engagement. Several emerging studies examine specific elements within the content that can serve as micro-level cues (e.g., contextual embedding, detailing, flattering) and macro-level cues (e.g., cohesive argument structuring), as well as meta-level cues (e.g., linguistic style matching) for detecting deception in the content [55,56]. For instance, based on speech act theory, deceivers tend to avoid providing detailed descriptions in their content [57]. Conversely, detailed content is considered sincere and credible, thus enhancing its perception by content consumers [58]. Therefore, we propose the following hypothesis:

H1a. *Content believability is positively correlated with user engagement on CCPs.*

Objectivity refers to the extent to which the content is unbiased and impartial on CCPs [32]. Within the realm of CCPs, content creators often subjectively express their own perspectives towards a specific entity through emotional sentiments. Maintaining objectivity on CCPs may not be conducive to attracting each form of user engagement. According to Heinonen [11], a like can be driven by entertainment purposes, with consumers investing minimal cognitive efforts in doing so [13,36]. By contrast, a comment represents an informative communication activity that requires more cognitive effort than a mere like [11,13,36]. Therefore, in brand communities on Facebook, Yang et al. [13] discovered that user-generated posts with positive sentiment receive more likes but fewer comments compared to those with neutral emotion. Moreover, they found that user-generated posts expressing negative sentiment only elicit likes rather than comments, confirming the existence of negative bias, referring to the phenomenon where people tend to give more weight to negative entities [59]. In comparison, the motivation behind a retweet is primarily driven by social connection and serves as an efficient mechanism for message dissemination with minimal cognitive requirements [11,15]. While both neutral and emotional content can be considered valuable information when evaluating IQ from a non-social interaction perspective, as highlighted by Gu and Konana [19], the dynamics change significantly when considering social interactions. Therefore, according to Stieglitz and Dang-Xuan [59], emotional messages are more likely to be retweeted quickly and frequently compared to those with neutral sentiment. Nevertheless, their findings do not support the existence of a negative bias. Therefore, we propose the following hypothesis:

H1b. *The positive impact of content objectivity on user engagement is more pronounced for comments compared to likes and retweets on CCPs.*

3.2. Contextual Quality and User Engagement

Contextual quality pertains to the perceived relevance within a given task [42,43], encompassing two primary dimensions of relevancy and informative richness in the context of CCPs following the research conducted by Cai et al. [21], Wang et al. [31], and Goh et al. [30].

From a consumer-centric perspective, relevancy refers to the value of content in fulfilling specific tasks for consumers [52,60]. In the context of CCPs, however, it is imperative to acknowledge that crowd content creation relies heavily on the creators' own abilities, motivation, and domain knowledge [22]. Consequently, the content remains agnostic to its use and caters to both known and potential future or unforeseen uses [22]. In other words, the information relevancy of content on CCPs is co-determined by content creators and consumers [61]. In this regard, CCPs commonly employ instance- or attribute-based data structures to provide consumers with sufficient flexibility in determining information usefulness. Therefore, we instead define the relevancy of content on CCPs as its value in offering references for consumers [56]. As for content consumers, they can compare the displayed instances or attributes with self-relevant information stored in memory through self-referencing, as defined in cognitive psychology [62]. Consequently, content created with a well-organized self-concept can enhance information receivers' elaboration and facilitate persuasive effects [55,61,62]. For instance, marketing content on social media platforms often openly conveys creators' personal experiences or inner thoughts to persuade content consumers [55]. However, the subjective nature of self-referential value in content requires careful consideration. Content consumers may express their appreciation or engage with it when they perceive its relevance to themselves; however, retweeting solely based on this aspect might be deemed irrelevant by the recipient. Therefore, we propose the following hypothesis:

H2a. *Content relevancy exhibits a positive correlation with likes and comments but does not significantly influence retweets on CCPs.*

The informative richness of the content on CCPs refers to the extent to which the content contains an adequate volume of information to fulfill specific requirements [1,52,60]. According to the S-O-R framework, incoming information can stimulate individuals' cognitive system in formulating a comprehensive perception of the content. Insufficient information may undermine receivers' perception of decision quality [63]. Particularly when communication takes place on CCPs, limitations on content length may result in an inadequate description of the focal entity, leading to heightened uncertainty and biased decision making [21]. Conversely, abundant information within the content can reduce readers' search costs for related instances from the crowd and provide them with a holistic cognitive representation of the focal entity [64]. In order to effectively convey their intentions, content creators on CCPs are increasingly incorporating visual elements, particularly photographs, into their content. Previous studies have extensively demonstrated the significance of photos in social media content as indicators of user engagement [28,65]. Consequently, Cai et al. [21] demonstrated the positive impact of informative richness, as measured by the word count of the content, on user engagement such as likes, comments, and shares received by a crowd-testing report. However, considering that retweets primarily serve information dissemination purposes, we propose the following hypothesis:

H2b. *The positive impact of informative richness of the content on user engagement is more pronounced for retweets compared to comments and likes on CCPs.*

3.3. Representational Quality and User Engagement

Representational quality refers to the level of clarity in presenting assessed information [42,43], which can be prominently exemplified by the readability within the content of CCPs, as indicated by Jin et al. [4] and Cai et al. [21]. Lee et al. [29] argue that directly informative content is associated with lower engagement on social media, implying that the amount of information within the content should be appropriate [32]. According to the effort–accuracy framework [66], due to limitations in information processing capacity, the final decision is made based on a compromise between making an accurate decision and minimizing cognitive effort. An excessive amount of information can exceed the cognitive load of users and result in suboptimal decision making. Lengthy expressions may pose comprehension challenges, whereas short ones could lack accuracy or completeness for users [60]. The presentation of comprehensive information should be concise and condensed; therefore, it is crucial to convey only the essential elements from the entirety of information. Consequently, readability measures the extent to which the content is effectively and comprehensively conveyed [52,60], which can reduce the cognitive efforts to be made before deciding to like, comment on, or retweet the content. Therefore, despite the varying cognitive input requirements for information processing across different forms of user engagement, we propose the following hypothesis:

H3. *Content readability is positively correlated with user engagement on CCPs.*

3.4. Source Credibility and User Engagement

Through a consumer-centric lens, the source credibility of information generally reflects a comprehensive subjective assessment derived from consumers' duration of usage [52,60]. According to signaling theory, reputation can serve as a signal to convey the intrinsic quality that may not be directly observable [67]. Drawing upon the ELM [46], reputation provides information receivers with a peripheral route that requires less cognitive effort and knowledge reserve in information processing for formulating their own attitudes and subsequent behaviors [42,43,48]. A positive reputation of both the content itself and its source has proven to be effective in enhancing consumers' perceptions of trustworthiness and encouraging their engagement [21,32]. A platform-validated reputation within the signaling environment of CCPs can serve as a credible endorsement for high-quality content [68] and is immune to manipulation [69]. In the context of e-commerce, a platform-validated reputation has largely proven to be associated with higher sales [70]. As reputational signals are typically conveyed through visual labels, such as stars and badges [71,72], they require minimal cognitive effort [47] and may subtly enhance the appeal of various forms of user engagement. Therefore, we propose the following hypotheses:

H4a. *The reputation of the content generated by the platform is positively correlated with user engagement on CCPs.*

H4b. *The reputation of the source generated by platform is positively correlated with user engagement on CCPs.*

4. Methodology

4.1. Data Collection

The data used for this study consist of travelogue content available on the MaBeeWoo platform (MaBeeWoo, https://www.mafengwo.cn/ accessed on 1 September 2024), the largest Chinese CCP dedicated to travelogues. This platform serves as a community where travelers openly share their extensive travel experiences, including dining options, lodging arrangements, transportation modes, sightseeing opportunities, shopping destinations, and recreational activities. The content is easily accessible to visitors without requiring any registration procedures. However, in order to utilize features such as liking, commenting, retweeting, or following someone on the travelogue page, visitors are required to log in using their phone number or email address. The MaBeeWoo dataset is highly suitable for estimating the impact of informational features on user engagement on CCPs due to four key reasons, as illustrated in Figure 2: (1) MaBeeWoo has minimal restrictions on content length but encourages visually appealing and detailed content through a star badge system provided by the platform. This ensures that sufficient informational features can be extracted. (2) Advertisements are rarely included in the travelogue content on MaBeeWoo, minimizing their influence on users' engagement decision making beyond the inherent characteristics of the content itself. (3) MaBeeWoo offers grade certification to content creators based on their contributions, prominently displayed alongside the content interface and accessible to users. (4) MaBeeWoo enables registered users to engage with content creators and other registered users through likes, comments, bookmarks, and retweets within a dedicated travelogue interface. It should be noted that content creators do not possess the ability to directly delete these records.

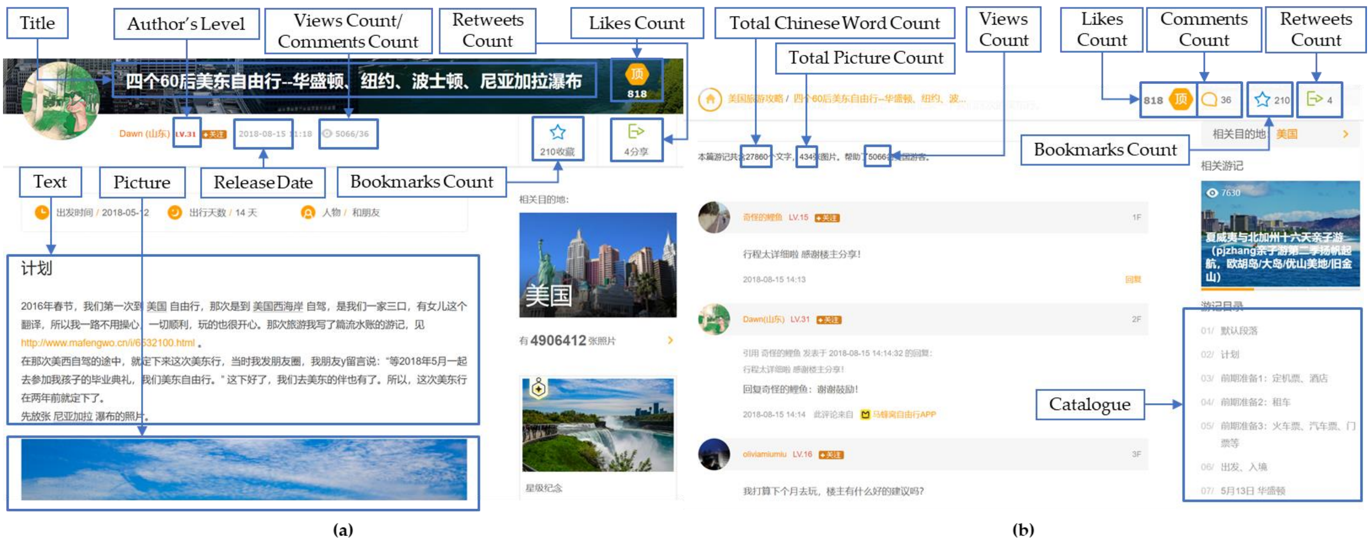

Figure 2. Screenshots of the travelogues available on the MaBeeWoo platform. Panel (**a**) displays the uppermost section of the page, providing comprehensive details regardingthe content creator and user engagement of the travelogue. Panel (**b**) showcases the lowermost section of the page, presenting statistical information pertaining to Chinese word count, picture count, and comments associated with the travelogue on this page.

In particular, we chose travelogues on the USA as a tourist destination from MaBeeWoo due to their abundance and well-organized nature, as these travel routes are typically expensive, time-consuming, and meticulously planned compared to domestic or neighboring outbound tours. We then adopted a step-by-step approach to extract key components from both the search results page and the travelogue interface. Specifically, we examined various attributes of each displayed travelogue on the search results page, including star badges (if available), title, author, release date, likes count, comments count, bookmarks count, and retweets count. Additionally, we used the URL links obtained in the previous step to locate each travelogue and thoroughly analyze its content on its respective homepage. This analysis included details such as author ranking, word count, and image count. Finally, data collection was conducted within a specific time frame spanning from 1 January 2016 to 31 December 2019. Consequently, we successfully crawled a total of 6956 travelogues.

4.2. Data Pre-Processing

Firstly, we implemented a three-step cleaning procedure to enhance data accuracy: (1) manually removing non-USA destinations; (2) uniquely identifying and eliminating duplicate travelogues based on URL field de-duplication in the database; and (3) manually deleting travelogues lacking essential fields (e.g., likes) or containing empty content. Following the data cleaning process, we obtained a final set of 4450 valid travelogues for subsequent analysis.

Subsequently, we proceeded with preprocessing the data to transform the unstructured travelogue text into a standardized and organized format, facilitating the extraction of informative features from each travelogue. By following the standard data preprocessing steps of deep learning models, this study pre-processed the cleaned travelogue data through word splitting and deactivation, word vector transformation, and corpus labeling [73]. Firstly, we utilized the Chinese word segmentation tool called Jieba for the purpose of word segmentation. Subsequently, based on two lexicons known as the "HIT deactivation word list" and "Baidu deactivation word list", we modified and adjusted the deactivation lexicon according to the specific data of this study. Following that, BERT (Bidirectional Encoder Representations from Transformers), a large-scale pre-trained model, was employed to transform each individual word into a vector within the vector space. Lastly, we approached named entity recognition and sentiment analysis as a sequence annotation problem, simultaneously extracting both entities and emotions using the joint extraction method [30].

As illustrated in Figure 2, MaBeeWoo enables nonexperts to create diverse content based on instances derived from their personal travel experiences by eliminating potential barriers to contribution, such as limitations in data input and crowd selection mechanisms [3]. In this context, a viable strategy for data standardization involves aligning the extracted entities and similar attributes [22]. Therefore, we used the six fundamental elements of tourism activities—food, accommodation, transportation, tours, shopping, and entertainment—to determine the attributes for annotating entities in travelogue content.

Regarding the emotion expressed in travelogue content, it was annotated as either positive or neutral since negative sentiments are not prominent in travelogues; considering them could have compromised prediction accuracy. We adopted the Chinese emotion ontology base proposed by the Dalian University of Technology. By utilizing the "BIO" annotation method, the annotation labels included five categories: B_pos (beginning of positive entity words), I_pos (middle word among positive entity words), B_norm (beginning of negative entity words), I_norm (middle word among negative entity words), and O (non-entity words). Table 3 provides an instance for the annotation in our dataset.

Table 3. An instance for annotation.

Sequence	Label
美	O
丽	O
的	O
纽	B_pos
约	I_pos

4.3. Model Design and Testing

The deep learning model we selected for extracting entities and their corresponding sentiment is the BERT+Bi-LSTM+CRF model, as illustrated in Figure 3. The Bi-LSTM (Bidirectional Long Short-Term Memory) model, widely employed in named entity recognition [74], combines forward and backward LSTM models to capture contextual semantic information. The CRF (Conditional Random Field algorithm) model leverages adjacent label information to obtain optimal values, complementing the Bi-LSTM model. Additionally, the BERT (Bidirectional Encoder Representations from Transformers) model, a pre-trained language model, effectively addresses word polysemy issues. Consequently, this combination of models yields an optimized outcome in deep learning by effectively handling large volumes of text entities, contextual relationships, and word polysemy encountered in travelogue content [75].

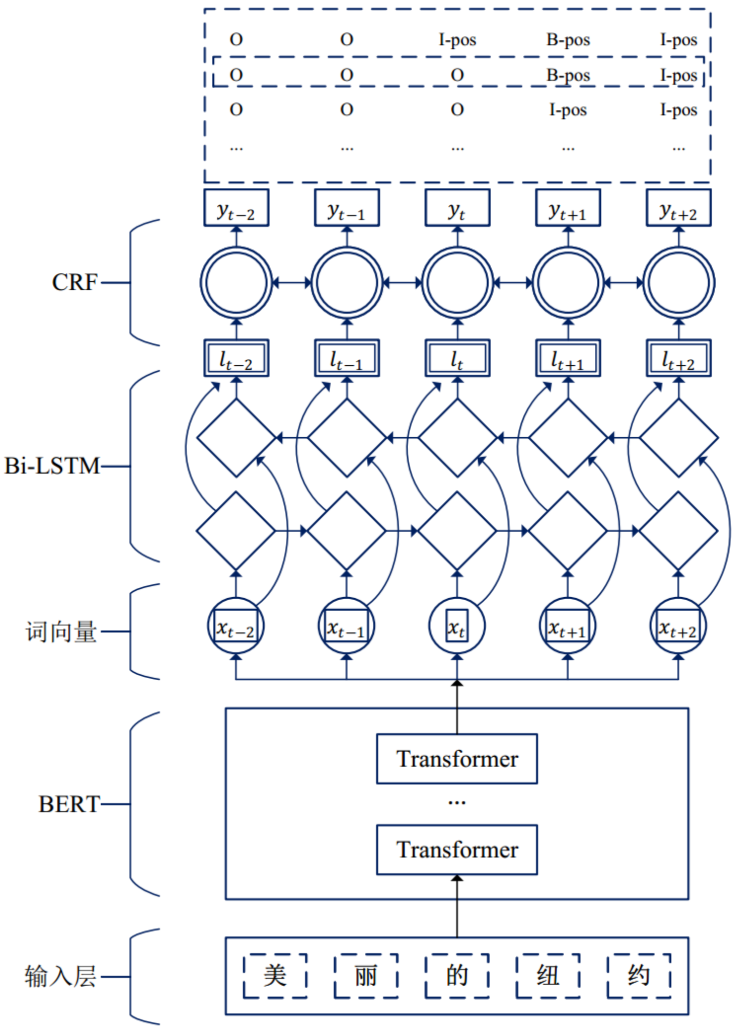

Figure 3. Framework diagram of the BERT+Bi-LSTM+CRF Model. The dashed line represents the input and output content, while the arrows indicate the flow of data and connections between neurons.The parameters were fine-tuned specifically for our dataset, based on the research conducted by Wang et al. [74] and Yue [75]. Specifically, we configured the Transformer layers to be 12, aligning with the BERT-base model. The maximum sequence length was set to 128, which corresponded to the longest string length in the dataset. Additionally, we set the BiLSTM hidden layer dimension as 128. After conducting multiple training runs, we determined that a batch size of 32 resulted in the most efficient convergence speed. In accordance with commonly used practices, we employed a learning rate of 0.001 for the Adam function. Finally, in line with default settings in BERT model architecture, we set the hidden layer dropout probability to 0.1.

The experimental dataset consisted of 500 travelogues randomly selected from the crawled travelogue dataset, which were manually labeled by a team comprising five undergraduate students and two graduate students. After assessing consistency using the kappa coefficient test, the data were further divided into training, testing, and validation sets in a ratio of 7:2:1 for model training. The precision, recall, and F1-score indexes indicating the effectiveness of the prediction are presented in Table 4. Finally, this study predicted the remaining travelogues with the trained model to obtain the entities and their corresponding sentiment.

Table 4. Experiment results of the model.

Category	Precision (P/%)	Recall (R/%)	F1-Score (%)
Entities with Neutral Sentiment	81.09	85.56	83.26
Entities with Positive Sentiment	75.09	79.35	77.16

4.4. Variable Measures

The measurement and descriptive statistics of the variables are presented in Table 5. The dependent variable was assessed through three distinct forms of user engagement, wherein likes represented users' production behavior for self-expression, comments indicated participation behavior for informative communication, and retweets measured message dissemination behavior for social interaction [11,35].

Table 5. The measurement of variables.

Dimension		Variable	Definition	Mean	S.D.	Min	Max
Dependent Variable	User Engagement	*Like*	The number of likes	85.789	206.856	0	2641
		Comment	The number of comments	7.027	22.220	0	368
		Retweet	The number of retweets	2.072	13.351	0	519
Central Route	Believability	*Detail*	The level of detail	0.0276	0.030	0	0.808
	Objectivity	*Neuemo*	The ratio of entities with neutral emotion	0.748	0.175	0	1
	Relevancy	*Refer*	The ratio of first-person pronouns	0.014	0.010	0	0.059
	Richness	*lnTextinf*	The volume of entities in the text	4.426	1.551	0	8.109
		lnPicinf	The volume of pictures	3.769	1.607	0	8.134
	Readability	*lnRead*	The reciprocal of textual information density	3.498	0.749	0.232	8.283
Peripheral Route	Content Reputation	*Contcert*	Whether or not there is a star badge	0.053	0.272	0	3
	Source Reputation	*lnSourcert*	The grade certification of the content creator	2.462	0.662	0	3.807
Controls		*Duration*	The release duration of the travelogue	7.178	0.299	6.594	7.661

The Central Route: Following the procedures of text mining and deep learning, we quantified certain informational features by calculating the ratio of specific words to the total number of words in the travelogue. The total Chinese word count was obtained using a word segmentation tool called Jieba. Subsequently, we evaluated the believability of the travelogue by analyzing the ratio of adjective words to all words (*Detail*) through dependency grammar analysis [55,56]. To assess its relevancy, we employed a commonly used indicator (*Refer*)—the ratio of first-person pronouns (e.g., I, we)—to all words [55,62]. We identified first-person pronouns in the segmented result and calculated their proportion. The objectivity of the travelogue was measured by extracting positive sentiment (assigned a value of 0) or neutral sentiment (assigned a value of 1) for each recognized entity through named entity recognition. These emotion values were then summed and divided by the total number of entities (*Neuemo*) in the travelogue [30].

The total number of these fundamental travel entities reflects the richness of textual information (*Textinf*) that users were most concerned about in the travelogue. Previous studies commonly employed the text length of the content on CCPs to reflect the information richness contained within [21,30], disregarding the nonsensical content that may dilute users' attention. Additionally, we utilized the number of pictures in the travelogue to gauge visual information richness ($Picinf$). Considering varying scales and sample bias, $Textinf$ and $Picinf$ were logarithmically transformed. The reciprocal of textual information density of the travelogue was employed as a measure of its readability (Read). A higher information density corresponds to a lower level of Readability in the travelogue. Information density was quantified by calculating the ratio between extracted textual entities and the total number of Chinese characters utilized in the travelogue.

The Peripheral Route: The platform's certification on both the travelogue and its creator was utilized as a reputational indicator for users to make informed engagement decisions [18,21,65,76]. MaBeeWoo awards star badges to high-quality travelogues, while no badge is assigned to those of lower quality. Consequently, we categorized low-quality travelogues as 0 and star-labeled ones as 1. Additionally, we assessed the reputation of creators based on the level of grade certification bestowed by MaBeeWoo.

Control Variables: Given the fact that an earlier publication of a travelogue yields amplified exposure and heightened user engagement in comparison to those with shorter publication durations [77], it was imperative to take into consideration the publication duration.

4.5. Model Specification

A travelogue remains static in terms of its content, while user engagement continues to evolve over time. Consequently, independent variables remain constant, whereas dependent variables experience substantial fluctuations. To tackle this problem, we employed cross-sectional data and included the duration in days since release as a control variable for further analysis. To alleviate possible multicollinearity issues among the variables, we also calculated correlation coefficients and variance inflation factors (VIFs). The findings confirmed no presence of multicollinearity among the variables as indicated by correlation coefficients below 0.6 and VIF values under 10.

According to the descriptive statistics in Table 3, both the number of likes and comments showed significantly higher levels of variability compared to their means, indicating substantial data dispersion. The likes and comments used in this study were non-negative count variables that deviated from the normal distribution. Therefore, we employed the negative binomial regression model to test the proposed hypotheses. The model was specified as Equation (1).

$$UE = \alpha + \beta CR + \gamma PR + \delta lnDuration + \varepsilon \quad (1)$$

$$UE = \alpha + \beta_1 Detail + \beta_2 Neuemo + \beta_3 Refer + \beta_4 lnTextinf + \beta_5 lnPicinf + \beta_6 lnRead + \gamma_1 Contcert + \gamma_2 lnSourcert + \varepsilon \quad (2)$$

The variable UE represents the level of user engagement measured by *Like*, *Comment*, or *Retweet*. CR is a set of variables related to content features that have a persuasive effect on user engagement through the central route. PR is a set of variables related to reputation that stimulate user engagement via the peripheral route. $lnDuration$ is the contral variable. ε indicates the interference item. Equation (1) is further specified as Equation (2) to estimate the impact of specific content features and reputation on user engagement.

5. Results

5.1. Model Results

To illustrate the specific associations between variables, we employed stepwise regression analysis to construct a model. Initially, control variables were incorporated into Model (1), followed by the gradual introduction of independent variables associated with both central and peripheral routes in Model (2) and Model (3). The econometric estimation results are presented in Table 6. The findings in Table 6 indicate consistent alignment among all four Models, suggesting robustness in the estimated results.

Regarding the intrinsic quality of travelogues, the results presented in Table 6 demonstrate a significant positive association between a higher level of detailed expression ($p = 0.000$, $\beta = 2.858$) and objectivity ($p = 0.000$, $\beta = 0.704$) with an increased number of likes. Similarly, travelogues that exhibit more comprehensive details ($p = 0.011$, $\beta = 1.755$) and objective tone ($p = 0.000$, $\beta = 0.943$) are more likely to attract a greater number of comments from content consumers. However, it is worth noting that both detailed expression ($p = 0.401$, $\beta = -1.405$) and objectivity ($p = 0.830$, $\beta = -0.084$) do not significantly impact the activity of retweeting by content consumers, thus partially supporting H1a. A comparison among the coefficients estimated for likes and comments reveals that comments display a stronger inclination towards objectivity in travelogues as hypothesized in H1b.

Table 6. Estimation results.

	Like				Comment				Retweet			
	(1)	**(2)**	**(3)**	**(4)**	**(1)**	**(2)**	**(3)**	**(4)**	**(1)**	**(2)**	**(3)**	**(4)**
lnDuration	0.793 ***	0.500 ***	1.063 ***	1.037 ***	1.706 ***	1.419 ***	2.577 ***	2.177 ***	2.387 ***	2.881 ***	4.075 ***	3.831 ***
	(0.136)	(0.127)	(0.094)	(0.085)	(0.229)	(0.207)	(0.143)	(0.138)	(0.510)	(0.343)	(0.248)	(0.242)
Detail		3.319 ***		2.858 ***		2.270 ***		1.755 **		−0.816		−1.405
		(0.424)		(0.354)		(0.877)		(0.689)		(1.604)		(1.673)
Neuemo		0.912 ***		0.704 ***		0.957 ***		0.943 ***		−0.361		−0.084
		(0.165)		(0.127)		(0.319)		(0.234)		(0.431)		(0.390)
Refer		27.118 ***		45.879 ***		4.479		11.666 ***		−11.372		−2.844
		(2.498)		(1.882)		(4.349)		(3.715)		(8.755)		(8.772)
lnTextinf		0.210 ***		0.203 ***		0.258 ***		0.240 ***		0.545 ***		0.412 ***
		(0.026)		(0.019)		(0.040)		(0.033)		(0.071)		(0.064)
lnPicinf		0.181 ***		0.041 **		0.330 ***		0.196 ***		0.239 ***		0.085 *
		(0.024)		(0.017)		(0.030)		(0.027)		(0.052)		(0.049)
lnRead		0.241 ***		0.198 ***		0.236 ***		0.210 ***		0.445 ***		0.304 ***
		(0.051)		(0.035)		(0.078)		(0.061)		(0.103)		(0.089)
Contcert			1.030 ***	0.855 ***			1.508 ***	1.078 ***			2.772 ***	2.062 ***
			(0.083)	(0.086)			(0.121)	(0.117)			(0.416)	(0.386)
lnSourcert			0.746 ***	0.981 ***			0.772 ***	0.714 ***			0.505 ***	0.453 ***
			(0.042)	(0.034)			(0.055)	(0.061)			(0.090)	(0.088)
Constant	−1.267	−2.952 ***	−5.348 ***	−8.921 ***	−10.416 ***	−12.720 ***	−19.003 ***	−19.572 ***	−16.664 ***	−25.174 ***	−30.797 ***	−32.133 ***
	(0.989)	(0.912)	(0.707)	(0.647)	(1.672)	(1.590)	(1.054)	(1.049)	(3.775)	(2.426)	(1.798)	(1.800)
pseudo R^2	0.003	0.023	0.033	0.061	0.013	0.052	0.061	0.084	0.022	0.076	0.086	0.108
chi2	33.859	460.277	640.317	2307.320	55.724	589.819	677.854	1233.172	21.941	502.181	357.861	559.023
N	4450	4450	4450	4450	4450	4450	4450	4450	4450	4450	4450	4450

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; values in parentheses are standard errors.

In terms of the contextual quality of travelogues, the findings presented in Table 5 demonstrate a significant positive correlation between the number of likes and extensive self-referencing ($p = 0.000$, $\beta = 45.879$), with higher levels of textual informativeness ($p = 0.000$, $\beta = 0.203$) and visual informativeness ($p = 0.018$, $\beta = 0.041$). Similarly, extensive self-referencing ($p = 0.002$, $\beta = 11.666$) and higher levels of textual ($p = 0.000$, $\beta = 0.240$) and visual ($p = 0.000$, $\beta = 0.196$) informativeness are more likely to attract a greater number of comments from content consumers. However, a significant positive correlation was only observed between the number of retweets and both textual informative richness ($p = 0.000$, $\beta = 0.412$) and visual informative richness ($p = 0.084$, $\beta = 0.085$). The self-referencing of the content for content consumers had no significant role ($p = 0.746$, $\beta = -2.844$) on retweets. These findings provide support for H2a. Furthermore, a stronger positive correlation was found between textual informativeness and retweets, while visual informativeness demonstrated a more prominent and positive association with comments, thus partially supporting H2b.

As for the representational quality of travelogues, there was an imperceptible positive relationship between readability and the number of likes ($p = 0.000$, $\beta = 0.198$), comments ($p = 0.001$, $\beta = 0.210$), and retweets ($p = 0.001$, $\beta = 0.304$). Notably, the coefficient for retweets was the largest, indicating that retweeting a travelogue demands higher readability compared to liking or commenting on it. Hypothesis 3 is supported.

Considering the source credibility of travelogues, both the endorsement of a star badge and grade certification received by content creators can significantly enhance user engagement in terms of likes, comments, and retweets. Notably, reputational signals that directly indicated IQ had a more pronounced impact on retweets ($p = 0.000$, $\beta = 2.062$) compared to likes ($p = 0.000$, $\beta = 0.855$) and comments ($p = 0.000$, $\beta = 1.078$). Conversely, reputational signals indicating the level of expertise of content creators had a more pronounced effect on likes ($p = 0.000$, $\beta = 0.981$) when compared to comments ($p = 0.000$, $\beta = 0.714$) and retweets ($p = 0.000$, $\beta = 0.453$). These findings support H4 and suggest the presence of distinct peripheral cues for encouraging different forms of user engagement.

5.2. Robustness Tests

To mitigate the potential impact of outliers on our results, this study conducted robustness checks by replacing the negative binomial regression model with an OLS model. Additionally, we logarithmically transformed the dependent variables (i.e., *Like*, *Comment*, and *Retweet*) to continuous variables. The regression analysis presented in Table 7 yielded consistent findings with those obtained from the negative binomial regression. In conclusion, our results are reliable and provide support for our hypotheses.

Table 7. The results estimated by using OLS.

	Like	Comment	Retweet
$lnDuration$	1.219 ***	1.730 ***	0.876 ***
	(0.056)	(0.042)	(0.037)
$Detail$	3.452 ***	1.151 *	−0.335
	(0.453)	(0.676)	(0.222)
$Neuemo$	0.708 ***	0.390 ***	0.004
	(0.097)	(0.075)	(0.050)
$Refer$	38.342 ***	4.782 ***	−1.168
	(1.836)	(1.337)	(0.974)
$lnTextinf$	0.171 ***	0.162 ***	0.107 ***
	(0.015)	(0.012)	(0.008)
$lnPicinf$	0.067 ***	0.095 ***	0.018 ***
	(0.013)	(0.010)	(0.007)
$lnRead$	0.187 ***	0.158 ***	0.064 ***
	(0.026)	(0.022)	(0.015)
$Contcert$	0.930 ***	1.010 ***	0.772 ***
	(0.059)	(0.062)	(0.062)
$lnSourcert$	0.684 ***	0.269 ***	0.162 ***
	(0.027)	(0.022)	(0.016)
Constant	−9.887 ***	−13.987 ***	−7.062 ***
	(0.432)	(0.321)	(0.278)
R^2	0.374	0.424	0.274
N	4450	4450	4450

Note: * $p < 0.1$, *** $p < 0.01$; values in parentheses are standard errors.

6. Discussions and Conclusions

With the proliferation of content consumption and entrepreneurship, CCPs have emerged as indispensable channels for individuals to access information. While user engagement is crucial for the growth of CCPs, platform sponsors still face the challenge of striking a balance between attracting diverse forms of user engagement due to their varied origins and functions in shaping value co-creation on CCPs. Effective guidance is required for content creators to align their content generation with the strategic interests of platform sponsors. To this end, this paper identifies the different determinantal attributes of IQ associated with three distinct forms of user engagement—likes, comments, and retweets. The summary of all hypothesis-based analyses is presented in Table 8. The results primarily demonstrate the following:

(1) Intrinsic IQ does not significantly influence retweets; however, it plays a crucial role in stimulating content consumers to engage through liking and commenting, suggesting a higher level of rationality associated with likes and comments compared to retweets, which may be driven by emotional sentiments.

(2) Among various quality attributes, relevancy exerts the most influential impact on stimulating engagement in terms of likes or comments, while exhibiting an insignificant effect on retweets. This implies a distinct set of factors that precede a retweet compared to likes and comments.
(3) Contrary to H2b, the visual informativeness of content was found to be more effective in eliciting comments rather than likes and retweets from content consumers. This finding can potentially be attributed to the congruence between images and textual content, as well as the affective and cognitive impressions evoked in consumers, which facilitate comment engagement by reducing the cognitive effort required [39,78].
(4) Representational quality and source credibility effectively promote all three forms of user engagement, as hypothesized. Particularly, content reputation has a dominant impact on retweets compared to other attributes related to IQ, indicating that retweeting is a more content-centric form of engagement on CCPs.

Table 8. Results of the hypotheses.

Hypotheses	Description	Results
H1a	Content believability is positively correlated with user engagement on CCPs.	Partially Supported
H1b	The positive impact of content objectivity on user engagement is more pronounced for comments compared to likes and retweets on CCPs.	Supported
H2a	Content relevancy exhibits a positive correlation with likes and comments but does not significantly influence retweets on CCPs.	Supported
H2b	The positive impact of informative richness of the content on user engagement is more pronounced for retweets compared to comments and likes on CCPs.	Partially Supported
H3	Content readability is positively correlated with user engagement on CCPs.	Supported
H4a	The reputation of platform-generated content is positively correlated with user engagement on CCPs.	Supported
H4b	The reputation of the source on a platform is positively correlated with user engagement on CCPs.	Supported

6.1. Theoretical Significance

Firstly, this paper comprehensively and conceptually distinguishes the three distinct forms of user engagement that can shape the dynamics of value co-creation on CCPs differently. Although previous studies have proposed typologies for consumer engagement behavior (CEB) based on either motivation or consumers' input from the perspective of Use and Gratification Theory [11,35,36] the most prevalent CEBs on CCPs, namely likes, comments, and retweets, have not been thoroughly aligned with these typologies and have continuously been treated as interchangeable measures of user engagement [21,29,40]. By considering user engagement behaviors as reflections of content consumers' emotional states and cognitive input [13], we provide a clear distinction between likes, comments, and retweets. The findings in this paper also reveal different antecedents for these distinct forms of user engagement, providing further evidence for their differentiation.

Secondly, this paper presents empirical evidence of the distinct factors influencing retweets compared to likes and comments. Yang et al. [13] differentiated between likes and comments based on the varying impact levels of discernible sentiment and topics in user-generated content on social media platforms. In our study, we further demonstrate that different attributes related to IQ, particularly believability, relevancy, and visual information richness, can significantly vary in their influence on likes and comments. However, from a holistic perspective, both likes and comments share similar antecedents in terms of IQ when compared to retweets. The aforementioned findings underscore the divergent patterns of information dissemination and production on CCPs, as emphasized by Park et al. [15], implying a distinct strategy design in the bidirectional diffusion of information on CCPs.

Thirdly, this paper presents a systematic approach for extracting IQ features from lengthy content on CCPs using deep learning algorithms. This enables us to validate the findings reported in previous studies in the literature regarding the IQ of short content in social media [55,58,59,65], providing a text analytic methodology for future research on CCPs that builds upon the work of Liu et al. [32]. The content on CCPs is generated through crowdsourcing, where the platform eliminates potential barriers to contribution such as data input limitations and crowd selection mechanisms, thereby facilitating content creation [3]. Accordingly, in accordance with the recommendation proposed by Lukyanenko et al. [22], we propose developing a conceptual model for content analysis based on domain knowledge. This entails utilizing a set of fundamental elements specific to the ontology of the content and subsequently aligning the extracted words with corresponding attributes within the conceptual model to perform named entity recognition. This allows us to assess both the textual information richness and readability of the content, as well as estimate their impact on user engagement.

6.2. Practical Significance

Firstly, this paper identifies information richness, readability, and source credibility as the most robust determinants for various forms of user engagement on CCPs, including likes, comments, and retweets. This implies that content creators and CCPs should prioritize considering the level of information richness embedded in their content while determining an optimal level of information density to avoid sacrificing the readability of said content. This aligns with the concept of "appropriate amount of information" [32] and "concise representation" [43] proposed in previous studies in the literature. Moreover, platforms should strategically design desirable criteria to guide content creation and subsequently stimulate the desired type of user engagement. By contrast, the insignificant role played by intrinsic quality and relevancy in retweets underscores the need for managing deceptive information diffusion and monitoring social connections among users to encourage valuable message dissemination from focal CCPs [23,55].

Secondly, CCPs with diverse strategic objectives should develop adaptable mechanisms that encourage specific aspects of IQ in order to account for the varying influence of IQ attributes on various types of user engagement. Given that likes serve as affirmative expressions promoting ongoing contribution [11–13], CCPs focusing on generating content ought to motivate creators to generate trustworthy and self-referential material in order to garner greater numbers of likes. Moreover, endorsements from platforms for content creators significantly contribute towards augmenting the quantity of received likes. On the other hand, comments can offer unique perspectives aimed at contributing to community development [11,14,35]. Therefore, CCPs that prioritize content diversity should focus on promoting objectivity and enhancing visual information richness to encourage a higher volume of comments. Lastly, CCPs emphasizing competitiveness and influence should strive for greater textual information richness and readability while actively endorsing the content to attract more retweets from consumers, as retweets can result in wider dissemination of the original content [11,15,37].

6.3. Limitations and Future Research Directions

Although this study yields some valuable conclusions, it still has the following limitations that should be addressed in further research. Firstly, our research is primarily based on travelogue data from MaBeeWoo, which shares some similar informational features with other CCPs such as Wikipedia, Stack Overflow, and Zhihu. However, it is important to note that our dataset may contain certain unique attributes specific to MaBeeWoo. Therefore, the findings of this study may not be universally applicable to all CCPs. Future research could estimate our findings in the context of other CCPs to enhance their generalizability. Secondly, this study employed binary variables (0–1) to quantify the positive or neutral sentiment associated with the entities prior to computing the overall sentiment of the travelogue; future research endeavors may consider assessing sentiment on a more nuanced, entity-specific level. Thirdly, although bookmarks are commonly used for user engagement in CCPs, their comprehension has been limited in prior studies, which hampers our inclusion of bookmarks within this paper. In future research, it would be beneficial to establish clear conceptual identification regarding bookmarks and thereby gain further insights into their antecedents concerning IQ.

Author Contributions: Y.W. conceptualized the idea, contributed to the study design, and wrote the first draft of the manuscript. M.W. ran the analyses and edited the original manuscript before submission. Z.Z. contributed to the research methodology and supervised this research. H.Y. organized the data collection. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Humanities and Social Science Foundation of the Ministry of Education of China, grant number 22YJAZH167.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zheng, Y.M.; Zhao, K.X.; Stylianou, A. The impacts of information quality and system quality on users' continuance intention in information-exchange virtual communities: An empirical investigation. *Decis. Support Syst.* **2013**, *56*, 513–524. [CrossRef]
2. Fader, P.S.; Winer, R.S. Introduction to the special issue on the emergence and impact of user-generated content. *Mark. Sci.* **2012**, *31*, 369–371. [CrossRef]
3. Lukyanenko, R.; Parsons, J.; Wiersma, Y.F.; Maddah, M. Expecting the unexpected: Effects of data collection design choices on the quality of crowdsourced user-generated content. *MIS Q.* **2019**, *43*, 623–647. [CrossRef]
4. Jin, Y.; Tan, Y.; Huang, J.H. Managing contributor performance in knowledge-sharing communities: A dynamic perspective. *Prod. Oper. Manag.* **2022**, *31*, 3945–3962. [CrossRef]
5. Zhao, K.X.; Zhang, P.Q.; Lee, H.M. Understanding the impacts of user- and marketer- generated content on free digital content consumption. *Decis. Support Syst.* **2022**, *154*, 113684. [CrossRef]
6. Wang, N.; Liu, Y.; Xiao, S.S. Which feedback matters? The role of expressions and valence in continuous high-quality knowledge contribution in the online Q&A community. *Decis. Support Syst.* **2022**, *156*, 113750.
7. Panico, C.; Cennamo, C. User preferences and strategic interactions in platform ecosystems. *Strateg. Manag. J.* **2022**, *43*, 507–529. [CrossRef]
8. Bapna, S.; Benner, M.J.; Qiu, L.F. Nurturing online communities: An empirical investigation. *MIS Q.* **2019**, *43*, 425–452. [CrossRef]
9. Kuang, L.N.; Huang, N.; Hong, Y.L.; Yan, Z.J. Spillover Effects of Financial Incentives on Non-Incentivized User Engagement: Evidence from an Online Knowledge Exchange Platform. *J. Manag. Inf. Syst.* **2019**, *36*, 289–320. [CrossRef]
10. Goes, P.B.; Guo, C.H.; Lin, M.F. Do incentive hierarchies induce user effort? evidence from an online knowledge exchange. *Inf. Syst. Res.* **2016**, *27*, 497–516. [CrossRef]
11. Heinonen, K. Consumer activity in social media: Managerial approaches to consumers' social media behavior. *J. Consum. Res.* **2011**, *10*, 356–364. [CrossRef]
12. Yan, Z.J.; Wang, T.; Chen, Y.; Zhang, H. Knowledge sharing in online health communities: A social exchange theory perspective. *Inform. Manag.* **2016**, *53*, 643–653. [CrossRef]
13. Yang, M.C.; Ren, Y.Q.; Adomavicius, G. Understanding user-generated content and customer engagement on Facebook business pages. *Inf. Syst. Res.* **2019**, *30*, 839–855. [CrossRef]
14. Zhang, Y.; Zhang, M.L.; Luo, N.; Wang, Y.; Niu, T. Understanding the formation mechanism of high-quality knowledge in social question and answer communities: A knowledge co-creation perspective. *Int. J. Inf. Manag.* **2019**, *48*, 72–84. [CrossRef]
15. Park, J.H.; Moon, J.Y.; Hong, S.J. Understanding the bi-directional message diffusion mechanism in the context of IT trends and current social issues. *Inform. Manag.* **2021**, *58*, 1–13. [CrossRef]
16. Malhotra, A.; Malhotra, C.K.; See, A. How to get your messages retweeted. *MIT Sloan Manag. Rev.* **2012**, *53*, 61–66.
17. Van Doorn, J.; Lemon, K.N.; Mittal, V.; Nass, S.; Pick, D.; Pirner, P.; Verhoef, P.C. Customer engagement behavior: Theoretical foundations and research directions. *J. Serv. Res.* **2010**, *13*, 253–266. [CrossRef]
18. Zhang, K.Z.K.; Barnes, S.J.; Zhao, S.J.; Zhang, H. Can consumers be persuaded on brand microblogs? An empirical study. *Inform. Manag.* **2018**, *55*, 1–15. [CrossRef]

19. Gu, B.; Konana, P. Competition among virtual communities and user valuation: The case of investing-related communities. *Inf. Syst. Res.* **2007**, *18*, 68–85. [CrossRef]
20. Yu, L.Y.; Chen, Z.X.; Yao, P.B.; Liu, H.D. A Study on the Factors Influencing Users' Online Knowledge Paying-Behavior Based on the UTAUT Model. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 1768–1790. [CrossRef]
21. Cai, J.X.; Ke, D.; Wu, J.; Luo, X.R. Get your report a thumb-up: An empirical investigation on crowd testing. *Decis. Support Syst.* **2022**, *158*, 113781. [CrossRef]
22. Lukyanenko, R.; Parsons, J.; Wiersma, Y.F. The IQ of the crowd: Understanding and improving information quality in structured user-generated content. *Inf. Syst. Res.* **2014**, *25*, 669–689. [CrossRef]
23. Wei, X.; Zhang, Z.; Zhang, M.Y.; Chen, W.Y.; Zeng, D.D. Combining crowd and machine intelligence to detect false news on social media. *MIS Q.* **2022**, *46*, 977–1008. [CrossRef]
24. Wang, R.Y.; Strong, D.M. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst.* **1996**, *12*, 5–33. [CrossRef]
25. Jia, J.Y.; Ma, G.F.; Li, H.Y.; Ding, J.Y.; Liu, K.N. Social power antecedents of knowledge sharing in project-oriented online communities. *IEEE Trans. Engin. Manag.* **2022**, *13*, 2493–2505. [CrossRef]
26. Yi, M.Y.; Yoon, J.J.; Davis, J.M.; Lee, T. Untangling the antecedents of initial trust in Web-based health information: The roles of argument quality, source expertise, and user perceptions of information quality and risk. *Decis. Support Syst.* **2013**, *55*, 284–295. [CrossRef]
27. Chen, J.V.; Su, B.C.; Widjaja, A.E. Facebook C2C social commerce: A study of online impulse buying. *Decis. Support Syst.* **2016**, *83*, 57–69. [CrossRef]
28. Karpenka, L.; Rudiene, E.; Morkunas, M.; Volkov, A. The Influence of a Brand's Visual Content on Consumer Trust in Social Media Community Groups. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 2424–2441. [CrossRef]
29. Lee, D.; Hosanagar, K.; Nair, H.S. Advertising content and consumer engagement on social media: Evidence from Facebook. *Manag. Sci.* **2018**, *64*, 5105–5131. [CrossRef]
30. Goh, K.Y.; Heng, C.S.; Lin, Z.J. Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content. *Inf. Syst. Res.* **2013**, *24*, 88–107. [CrossRef]
31. Wang, X.L.; Lu, J.D.; Ow, T.T.; Feng, Y.Q.; Liu, L.N. Understanding the emotional and informational influence on customer knowledge contribution through quantitative content analysis. *Inform. Manag.* **2021**, *58*, 103426. [CrossRef]
32. Liu, X.M.; Wang, G.A.; Fan, W.G.; Zhang, Z.J. Finding useful solutions in online knowledge communities: A theory-driven design and multilevel analysis. *Inf. Syst. Res.* **2020**, *31*, 731–752. [CrossRef]
33. Dolan, R.; Conduit, J.; Fahy, J.; Goodman, S. Social media engagement behaviour: A uses and gratifications perspective. *J. Strateg. Mark.* **2016**, *24*, 261–277. [CrossRef]
34. Oh, C.; Roumani, Y.; Nwankpa, J.K.; Hue, H.F. Beyond likes and tweets: Consumer engagement behavior and movie box office in social media. *Inform. Manag.* **2017**, *54*, 25–37. [CrossRef]
35. Shao, G.S. Understanding the appeal of user-generated media: A uses and gratification perspective. *Inf. Res.* **2009**, *19*, 7–25. [CrossRef]
36. Courtois, C.; Mechant, P.; De Marez, L.; Verleye, G. Gratifications and seeding behavior of online adolescents. *J. Comput.-Mediat. Commun.* **2009**, *15*, 109–137. [CrossRef]
37. Shi, Z.; Rui, H.X.; Whinston, A.B. Content sharing in a social broadcasting environment: Evidence from twitter. *MIS Q.* **2014**, *38*, 123–142. [CrossRef]
38. Brodie, R.J.; Ilic, A.; Juric, B.; Hollebeek, L. Consumer engagement in a virtual brand community: An exploratory analysis. *J. Busi. Res.* **2013**, *66*, 105–114. [CrossRef]
39. He, Z.; Deng, N.; Li, X.R.; Gu, H. How to "read" a destination from images? machine learning and network methods for DMOs' image projection and photo evaluation. *J. Travel Res.* **2022**, *61*, 597–619. [CrossRef]
40. Hwong, Y.L.; Oliver, C.; Van, K.M.; Sammut, C.; Seroussi, Y. What makes you tick? The psychology of social media engagement in space science communication. *Comput. Hum. Behav.* **2017**, *68*, 480–492. [CrossRef]
41. DeLone, W.H.; McLean, E.R. The DeLone and McLean model of information systems success: A ten-year update. *J. Manag. Inf. Syst.* **2003**, *19*, 9–30.
42. DeLone, W.H.; McLean, E.R. Information systems success: The quest for the dependent variable. *Inf. Syst. Res.* **1992**, *3*, 60–95. [CrossRef]
43. Ghasemaghaei, M.; Hassanein, K. Online information quality and consumer satisfaction: The moderating roles of contextual factors—A meta-analysis. *Inform. Manag.* **2015**, *52*, 965–981. [CrossRef]
44. Lee, Y.W.; Strong, W.; Kahn, B.K.; Wang, R.Y. AIMQ: A methodology for information quality assessment. *Inform. Manag.* **2002**, *40*, 133–146. [CrossRef]
45. Yoo, C.W.; Huang, C.D.; Goo, J. Task support of electronic patient care report (ePCR) systems in emergency medical services: An elaboration likelihood model lens. *Inform. Manag.* **2020**, *57*, 103336. [CrossRef]
46. Petty, R.E.; Cacioppo, J.T.; Petty, R.E.; Cacioppo, J.T. The elaboration likelihood model of persuasion. In *Advances in Experimental Social Psychology*; Academic Press: San Diego, CA, USA, 1986.

47. Kitchen, P.J.; Kerr, G.; Schultz, D.E.; McColl, R.; Pals, H. The elaboration likelihood model: Review, critique and research agenda. *Euro. J. Mark.* **2014**, *48*, 2033–2050. [CrossRef]
48. Petty, R.; Cacioppo, J. Central and Peripheral Routes to Persuasion: Application to Advertising. *Advert. Consum. Psychol.* **1983**, *1*, 3–23.
49. Petty, R.E.; Kasmer, J.A.; Haugtvedt, C.P.; Cacioppo, J.T. Source and message factors in persuasion: A reply to Stiff's critique of the elaboration likelihood model. *Commun. Monographs.* **1987**, *54*, 233–249. [CrossRef]
50. Coulter, K.S.; Girish, N.P. The effects of cognitive resource requirements, availability, and argument quality on brand attitudes: A melding of elaboration likelihood and cognitive resource matching theories. *J. Advert.* **2004**, *33*, 53–64. [CrossRef]
51. Petty, R.E.; Wegener, D.T.; Fabrigar, L.R. Attitudes and attitude change. *Annual Rev. Psychol.* **1997**, *48*, 609–647. [CrossRef]
52. Agarwal, N.; Yiliyasi, Y. *Information Quality Challenges in Social Media*; ICIQ: Tarragona, Spain, 2010.
53. Jensen, M.L.; Averbeck, J.M.; Zhang, Z.; Wright, K.B. Credibility of anonymous online product reviews: A language expectancy perspective. *J. Manag. Inf. Syst.* **2013**, *30*, 293–323. [CrossRef]
54. Hancock, J.T.; Curry, L.E.; Goorha, S.; Woodworth, M. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes.* **2007**, *45*, 1–23. [CrossRef]
55. Wang, F.; Xu, H.; Hou, R.; Zhu, Z. Designing marketing content for social commerce to drive consumer purchase behaviors: A perspective from speech act theory. *J. Retail. Consum. Serv.* **2023**, *70*, 103156. [CrossRef]
56. Ludwig, S.; Van Laer, T.; De Ruyter, K.; Friedman, M. Untangling a Web of lies: Exploring automated detection of deception in computer-mediated communication. *J Manag Inf Syst.* **2016**, *33*, 511–541. [CrossRef]
57. DePaulo, B.M.; Lindsay, J.J.; Malone, B.E.; Muhlenbruck, L.; Charlton, K.; Cooper, H. Cues to deception. *Psychol. Bulletin.* **2003**, *129*, 74–118. [CrossRef]
58. Ansari, S.; Gupta, S. Customer perception of the deceptiveness of online product reviews: A speech act theory perspective. *Int. J. Inf. Manag.* **2021**, *57*, 14. [CrossRef]
59. Stieglitz, S.; Dang-Xuan, L. Emotions and information diffusion in social media-sentiment of Microblogs and sharing behavior. *J. Manag. Inf. Syst.* **2013**, *29*, 217–247. [CrossRef]
60. Fisher, C.; Lauría, E.; Chengalur-Smith, S.; Wang, R. *Introduction to Information Quality*; MIT Inf Quality Program: Cambridge, MA, USA, 2008.
61. Wang, H.S.; Noble, C.H.; Dahl, D.W.; Park, S. Successfully communicating a cocreated innovation. *J. Mark.* **2019**, *83*, 38–57. [CrossRef]
62. Escalas, J.E. Self-referencing and persuasion: Narrative transportation versus analytical elaboration. *J. Consum. Res.* **2007**, *33*, 421–429. [CrossRef]
63. Xu, J.J.; Benbasat, I.; Cenfetelli, R.T. The nature and consequences of trade-off transparency in the context of recommendation agents. *MIS Q.* **2014**, *38*, 379–406. [CrossRef]
64. Kim, S.E.; Lee, K.Y.; Shin, S.I.; Yang, S.B. Effects of tourism information quality in social media on destination image formation: The case of Sina Weibo. *Inform. Manag.* **2017**, *54*, 687–702. [CrossRef]
65. Jin, S.V.; Ryu, E. "I'll buy what she's #wearing": The roles of envy toward and parasocial interaction with influencers in Instagram celebrity-based brand endorsement and social commerce. *J. Retail. Consum. Serv.* **2020**, *55*, 102121.
66. Payne, J.W.; Bettman, J.R.; Johnson, E.J. *The Adaptive Decision Maker*; Cambridge University Press: Cambridge, UK, 1993.
67. Connelly, B.L.; Certo, S.T.; Ireland, R.D.; Reutzel, R.C. Signaling theory: A review and assessment. *J. Manag.* **2011**, *37*, 39–67. [CrossRef]
68. Hukal, P.; Henfridsson, O.; Shaikh, M.; Parker, G.G. Platform signaling for generating platform content. *MIS Q.* **2020**, *44*, 1177–1205. [CrossRef]
69. Nakayama, M.; Wan, Y. Exploratory Study on Anchoring: Fake Vote Counts in Consumer Reviews Affect Judgments of Information Quality. *J. Theor. Appl. Electron. Commer. Res.* **2017**, *12*, 1–20. [CrossRef]
70. Zhang, S.Y.; Lee, D.; Singh, P.V.; Srinivasan, K. What makes a good image? Airbnb demand analytics leveraging interpretable image features. *Manag. Sci.* **2021**, *68*, 5644–5666. [CrossRef]
71. Fisher, M.; Newman, G.E.; Dhar, R. Seeing stars: How the Binary Bias Distorts the Interpretation of Customer Ratings. *J. Consum. Res.* **2018**, *45*, 471–489. [CrossRef]
72. Cheng, H.K.; Fan, W.G.; Guo, P.P.; Huang, H.L.; Qiu, L.F. Can "Gold Medal" Online Sellers Earn Gold? The Impact of Reputation Badges on Sales. *J. Manag. Inf. Syst.* **2020**, *37*, 1099–1127. [CrossRef]
73. Colladon, A.F.; Guardabascio, B.; Innarella, R. Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. *Decis. Support Syst.* **2019**, *123*, 113075. [CrossRef]
74. Wang, Y.; Currim, F.; Ram, S. Deep Learning of Spatiotemporal Patterns for Urban Mobility Prediction Using Big Data. *Inf. Syst. Res.* **2022**, *33*, 579–598. [CrossRef]
75. Yue, Z.Y.T.J.Z. Algorithm Term Extraction and Innovation Evolution Path Construction Based on BERT BiLSTM-CRF Model. *Inf. Sci.* **2022**, *40*, 71–78.
76. Cheng, X.S.; Gu, Y.; Hua, T.; Luo, X.R. The paradox of word-of-mouth in social commerce: Exploring the juxtaposed impacts of source credibility and information quality on SWOM spreading. *Inform. Manag.* **2021**, *58*, 11. [CrossRef]

77. Zhang, Z.Q.; Zhang, Z.L.; Chen, P.Y. Early bird versus late owl: An empirical investigation of individual shopping time habit and its effects. *MIS Q.* **2021**, *45*, 117–162. [CrossRef]
78. Ceylan, G.; Diehl, K.; Proserpio, D. Words meet photos: When and why photos increase review helpfulness. *J. Mark. Res.* **2023**, *61*, 5–26. [CrossRef]