

Content Moderation Effect: How to Build the Video Influence of Bilibili Through Bullet Screen Interaction Types

Licheng Wang

*School of Hotel and Tourism Management, Macau University of Science and Technology, Macau,
China*

v01d23123@outlook.com

Abstract. Although the theory of the interactive ritual chain can explain the cohesion of online communities, it is difficult to explain the differences in ritual effectiveness due to the neglect of interactive content. Therefore, this study focuses on the bullet comments in the knowledge area of Bilibili and proposes the "content moderation effect" as the core explanatory mechanism. Through content analysis and fixed effects model regression of 45326 bullet comments from 300 videos, this study constructed and validated a typological framework. Research has found that the long-term knowledge value of content interactive barrage driven videos (collection rate $\beta=0.32$, $p<0.001$); And creative memes and bullet comments stimulate users' immediate feedback behavior (like rate $\beta=0.24$, $p<0.001$). This study suggests that the efficacy of rituals depends not only on structural elements, but also on the systematic regulation of the flow of symbols and information quality within them. This discovery not only revolutionizes the theory of digital interaction, but also challenges the regulatory power of platform algorithms. The micro findings of this study touch upon a core tension of the digital age: under the wave of "quantifying everything", human emotions and cognition are at risk of being simplified into data.

Keywords: Bullet chat, interactive ritual chain, typology, content moderation effect, Bilibili

1. Introduction

The core contradiction in bullet screen research lies in the disconnect between highly complex platform interaction practices and overly simplified existing theoretical explanations. This is specifically manifested as a double theoretical blind spot: research or falling into the 'quantity paradox' that only focuses on the scale of interaction; or is limited by the incompleteness of the explanation of the interactive ritual chain theory, which is better at diagnosing the existence of "rituals", but neglects to explain the differences in "ritual efficacy". The true impact mechanism of the barrage on video dissemination is still a "black box" due to the combined effect of both factors.

To test the 'content moderation effect'(i.e. the systematic regulation of ritual efficacy by the quality of interactive content), This study designed a two-stage empirical plan: Firstly, reveal the type spectrum of barrage interactions through a content analysis system (clarify "what is regulating"); Subsequently, a fixed effects model is used to quantify the differential impact of different types on communication effects (measuring the "moderation effect").

Based on this, this study aims to address three progressive questions:(1) How to systematically construct the barrage typology of Bilibili knowledge area? (2) How do different types of bullet comments drive instant feedback (likes), substantial incentives (coins), and long-term value (collections) respectively? (3) How can the above findings jointly promote the theory of interactive ritual chain and even broader research on digital interaction, achieving a visual shift from structural analysis to content regulation?

Therefore, this study aims to address the theoretical dilemma mentioned above, with its core contribution reflected in the dual advancement of theory and application: At the theoretical level, by establishing the 'content moderation effect', Provide key micro regulatory mechanisms for the theory of interactive ritual chains, Thus systematically promoting the necessary transformation of the theory from a structural paradigm to a content paradigm; At the practical level, research has found that it can be directly transformed into a barrage scene planning strategy based on the framework of "value anchor points" and "emotional high points", provide a practical decision-making framework rooted in empirical evidence for content creation and platform governance.

2. Literature review and theoretical framework

2.1. Theoretical blind spots and "content shift" in interactive ritual chains

Although Collins' theory of interactive ritual chains accurately depicts the structural elements of ritual formation, but its classic assumption regards ritual output (such as emotional energy) as a homogeneous result, but its classic assumption regards ritual output (such as emotional energy) as a homogeneous result, failed to explain the systematic regulation of the final efficacy intensity by the differences in the quality and type of symbol content [1]. To analyze this core mechanism, this study proposes the "content moderation effect", advocating that ritual efficacy not only depends on structural integrity, but is also crucially regulated by the information quality and symbolic attributes of interactive content.

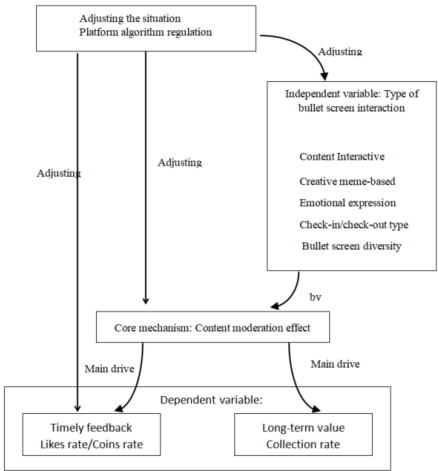


Figure 1. Research conceptual framework diagram

This perspective resonates deeply with the "social signal amplification" mechanism proposed by Chen et al., which emphasizes the differential shaping of collective response intensity by specific content characteristics (such as emotional concentration and community consensus) [2]. This mechanism indicates that such content can be selectively amplified through algorithmic

recommendations and user imitation paths, which is directly consistent with the findings of this study on the efficient catalysis of instant emotional energy in "creative memes" bullet comments.

In summary, this study, together with relevant cutting-edge advances, demonstrates the necessity and feasibility of implementing a "content shift" in the theory of interactive ritual chains, shifting the focus of analysis from "whether the ritual exists" to "how the content in the ritual specifically regulates its effectiveness" [3]. Caption: Social interaction type barrage has no statistically observable significant impact on video dissemination effect (As shown in Figure 1).

2.2. Information ecological chain: from role positioning to efficiency quantification

The theory of information ecological chain defines barrage users as "information decomposers", referring to their active processing of video information through behaviors such as supplementation, correction, and questioning. However, existing studies have failed to reveal the efficacy differences of different behavioral patterns in collaborative processing of such information [4]. To this end, this study constructs a barrage typology framework, operationalizes abstract roles into specific interactive categories, and uses regression analysis to clarify their efficacy ranking (such as the prominent driving effect of "knowledge supplementation" on long-term collection value). This achieves a key advancement in theory from macro role description to micro efficiency quantification, providing a detailed explanation for understanding the structural differences in user contributions in knowledge communities.

2.3. Effect measurement: building a multidimensional framework

On the basis of quantifying the effectiveness of interactive behavior, in order to surpass the limitations of traditional playback metrics and achieve precise measurement of user commitment gradients, this study constructs a three-dimensional communication effect framework: Instant recognition (like rate), substantial incentives (coin insertion rate), and long-term value (collection rate) correspond to emotional feedback, value identification, and knowledge storage expectations, respectively [5]. This framework advances the measurement of effects from the judgment of "presence or absence" to the analysis of "nature and depth of action", thereby accurately testing whether there is a differentiated path in the impact of bullet screen interaction on the above dimensions.

3. Research methods

3.1. Data collection and sampling

This study used stratified systematic sampling to randomly select 300 videos from eight sub sections of the Bilibili knowledge area based on the creator's fan base (<100000, 100000-500000, >500000) to ensure their representativeness [6]. The research data is obtained through the public API of Bilibili, including video metadata and full barrage.

3.2. Construction and verification of barrage typology framework

Based on a preliminary analysis of 45326 bullet comments and integrating symbolic interactionism and information ecological chain theory, this study constructed a typological framework as shown in Table 1 [4,7]. The framework has passed double-blind coding test (Cohen Kappa coefficient=0.81) and has good reliability.

Table 1. Typological framework for barrage interaction

category	subcategory	Description and Examples	theoretical origins
Interactive content category	Knowledge supplementation	Supplement video information or correct errors	Theory of Information Ecological Chain [4]
	Ask questions and doubts	Ask questions about the video content	Theory of Information Ecological Chain [4]
	Understand and Confirm	Express that you have understood	Theory of Information Ecological Chain [4]
Creative memes	Internal meme	Memes created around UP hosts or video content	Interactive Ritual Chain Theory [1]
	External meme	Using internet slang for re creation	Interactive Ritual Chain Theory [1]
Emotional expression category	Positive emotions	Expressing love and appreciation	Interactive Ritual Chain Theory [1]
	Negative emotions	Expressing dissatisfaction and boredom	Interactive Ritual Chain Theory [1]
	Surprise expression	Expressing shock and emotion	Interactive Ritual Chain Theory [1]
Check in category	Simple check-in	Declare their own existence	Symbolic interaction theory [7]
	Watch history	Record the number of views or progress	Symbolic interaction theory [7]
Social interaction category	Reply to others	Directly respond to barrage comments from other users	Social Presence Theory [8]
	atmosphere creation	Call for collective action and create a sense of ceremony	Social Presence Theory [8]

3.3. Variable measurement and identification strategies

The operationalized measurement system used in this study strictly follows the principle of causal inference to maximize internal validity, and is used to empirically test the theoretical model.

The dependent variable (video dissemination effect) starts from the depth of user cognition and engagement, and is operationalized into three observation indicators with theoretical gradients: like rate (number of likes/views) measures instant, low threshold emotional resonance; The value recognition and substantive support of measuring the depth of coin insertion rate (number of coins inserted/number of views); Collection rate (number of collections/views) measures long-term knowledge storage and future consumption intention. This multidimensional system makes it possible to accurately separate the differentiated impact paths of barrage interactions on immediate feedback, substantive incentives, and long-term value.

The core independent variable (barrage structure characteristics) directly serves the testing of core research hypotheses: (1) The proportion of barrage types (calculated based on the typology framework in Table 1) is used to verify the functional efficacy differences of different interaction modes, namely the existence and intensity of the "content moderation effect"; (2) The diversity index of barrage (evaluated by type richness and distribution uniformity) is used to verify the independent gain effect of the diversity of barrage ecology on the dissemination effect.

The control variables aim to isolate three key sources of confusion in the system: The video duration isolates the interference of the complexity and investment cost of the content itself; The

effect of isolating creators based on the number of UP main fans has both reputation and traffic foundation; The trend of accumulated isolation time and natural exposure in video release duration. This control strategy together forms a rigorous identification framework to strip away competitive explanations and ultimately clearly identify the net effect of barrage interactions on communication effectiveness.

To control unobservable endogeneity issues, a video individual fixed effects model is adopted:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma C_{it} + \mu_i + \varepsilon_{it} \tag{1}$$

Among them, μ_i is the fixed effect of individual video, used to control the inherent features of the video that do not change over time [9].

This study used the instrumental variable method for robustness testing, with the "proportion of similar bullet comments in adjacent videos" as the instrumental variable (as shown in Figure 2). The validity of this variable is based on: (1) Relevance, as algorithm recommendations are similar to user groups, adjacent video barrage styles have spatial correlation; (2) Exogeneity, after controlling for individual fixed effects (μ_i), mainly affects the target video through shaping community interaction norms (imitation effect), rather than directly affecting the dissemination effect through other unobserved pathways. The weak instrumental variable test supports its validity (Cragg Donald Wald F statistic=19.83>16.38).

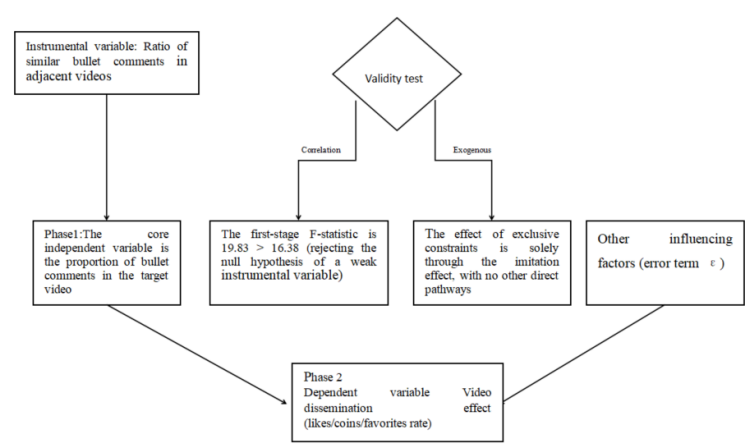


Figure 2. Logical framework and effectiveness test of instrumental variable method

4. Research results

4.1. Differentiation of bullet screen types and dissemination effects

The fixed effects regression results reveal a clear functional differentiation mechanism, as shown in Table 2 below.

Table 2. Regression analysis results of barrage types and dissemination effects (standardized coefficients)

Bullet chat type	like rate	Coin insertion rate	collection rate
Check in category	-0.08	-0.12	-0.15
Emotional expression category	0.21	0.17	0.09
Interactive content category	0.18	0.12	0.32
Creative memes	0.24	0.25	0.11
Social interaction category	0.05	0.08	0.07
Diversity of bullet screens	0.19	0.16	0.27

Note: The model includes video fixed effects and a complete set of control variables. $p < .05, p < .01, p < .001$

The table data shows that creative memes have the most significant driving effect on instant feedback, which stems from their unique advantages as internal symbols in quickly building group unity and emotional energy; Content interactive barrage is the value storage of videos, which drives user collection behavior through knowledge collaborative production. It has the strongest positive correlation with collection rate and significantly enhances the long-term reference value of content; The high-density punch in bullet comments weaken the perceived value of the content through a "signal dilution" mechanism.

The instrumental variable estimation is highly consistent with the results of the fixed effects model, confirming the robustness of the core findings. The sample regression further shows that the positive effect of interactive bullet comments is particularly significant among middle and tail end creators, revealing that content creators at different growth stages have differentiated interactive strategy needs.

5. Discussion

5.1. Moving towards a qualitative paradigm of interactive research

This study has three core advances in digital interaction theory. Firstly, promote the "content shift" of the interactive ritual chain theory. Research has found that emotional energy in Collins' theory is not homogenized output, but is systematically regulated by the flow of symbols and information quality in rituals. This' content moderation effect 'accurately identifies the core blind spot of the original theory in explaining differences in interactive effectiveness. Secondly, deconstruct the illusion of participation. The proliferation of low-quality bullet comments can release negative signals of "shallow content", thereby eroding the surface value of high mutual momentum [10]. This discovery reveals, from an empirical perspective, the inherent contradictions in platforms that solely pursue interactive scale. Thirdly, construct an integrated framework of "functional complementarity". Content interaction (knowledge enhancement) and creative memes (emotional cohesion) work together through a collaborative mechanism to optimize the dissemination effect. This framework not only forms a dialogue with the theory of media richness, but also promotes the substantive integration of theory by introducing micro mechanisms based on interactive content, providing a systematic explanation for qualitative dimension interaction research [11].

5.2. Practical significance and critical reflection

Based on the "content moderation effect", this study proposes a "barrage scene planning" framework to guide creators in accurately designing interactions: setting value anchors in knowledge difficulties (such as open-ended questions or common misconceptions) to stimulate in-depth discussions; Pre embed emotional highlights at interesting points (such as "internal memes" or resonance symbols), promote creative interaction and collective resonance, systematically guide high-quality bullet comments, and optimize the long-term and short-term value of content.

This effect has a dual nature in the platform context: it is both an optimization tool for creators and a means of algorithm training. Once the platform defines "high-value" bullet comments through quantitative rewards, users' spontaneous participation is easily transformed into "strategic performances" chasing traffic, highlighting the sustained tension between user agency and platform regulation under algorithmic governance [12].

6. Conclusion

This study empirically reveals the differential impact mechanism of barrage interaction on video dissemination effects by constructing a typology framework. The core conclusion is that barrage interaction is not homogeneous background noise, but a strategic resource with clear type structure and functional differentiation. The core breakthrough of this study lies in the empirical establishment of the "content moderation effect", which promotes the "content shift" of the interactive ritual chain theory; And by integrating fixed effects models and instrumental variable methods, it provides a more rigorous empirical path for causal inference of community interaction.

This study also has three limitations: In terms of external validity, the sample is concentrated in the knowledge area of Bilibili, and the universality of the conclusion in areas such as entertainment and life needs to be tested; In terms of measurement validity, bullet screen quality measurement can be further deepened by combining natural language processing technology; In causal inference, it is mainly based on observational data, and supplementary evidence can be provided through controlled field experiments in the future.

In the context of knowledge payment and the rise of online education, understanding user interaction has become the key to digital content competitiveness. This study not only provides empirical evidence for this, but also reveals a fundamental contradiction in the digital age: while the data logic of "quantifying everything" expands the possibilities of interaction, it systematically creates the risk of emotional alienation by indicators. Future research needs to focus on how users engage in creative practices in micro interactions such as bullet comments, seek meaning reconstruction in algorithmic regulation, and explore the authenticity of digital communication.

References

- [1] Collins, R. (2004) *Interaction Ritual Chains*. Princeton University Press.
- [2] Chen, Y., Wang, L. and Zhang, G. (2024) Social Signal Amplification in Digital Communities. *Journal of Computer-Mediated Communication*.
- [3] Li, X.Y. (2018) Research on Barrage Culture from the Perspective of Symbolic Interaction Theory. *Today's Media*.
- [4] Zhou, T., Liu, R.J. and Zhang, J.T. (2023) Research on the Composition and Evolutionary Mechanism of the Information Ecological Chain in Online Knowledge Communities. *Intelligence Theory and Practice*, 46(5), 45–52.
- [5] Zhou, B.H. (2021) Quantitative and Qualitative: A Two-Dimensional Perspective on Understanding Communication Effects in the Digital Age. *Journalism University*, (9), 1–12.
- [6] Babbie, E.R. (2021) *The Practice of Social Research* (15th ed.). Cengage Learning.

- [7] Wang, Y.Q. and Rong, M.D. (2024) Research on Short Video Barrage Interaction from the Perspective of Interaction Ritual Chain Theory: A Case Study of Bilibili Couple Videos. *News and Communication Review*, 77(1), 56–70.
- [8] National Taiwan Normal University. (2024) The Relationship Between Bullet Screen Function Quality and Community Loyalty: The Mediating Effect of Social Presence and Playfulness. Unpublished Master's Thesis, Department of Social Education.
- [9] Wooldridge, J.M. (2016) *Introductory Econometrics: A Modern Approach* (6th ed.). Cengage Learning.
- [10] Spence, M. (1973) Job Market Signaling. *The Quarterly Journal of Economics*.
- [11] Daft, R.L. and Lengel, R.H. (1986) Organizational Information Requirements, Media Richness and Structural Design. *Management Science*.
- [12] Gillespie, T. (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.