

Construction and Case Analysis of a Cocooning Degree Measurement Model for Online New Media Information

Yanning Yang*

School of No.80 High School of Beijing, Beijing 100102, China

*bbmmyy_cn@163.com

Abstract

In the era of online new media, algorithmic recommendations are widely employed, giving rise to the phenomenon of information silos. This has adverse effects on user cognition and the online ecosystem, making quantitative research on this topic of significant practical importance. This study collected user behaviour and content data from multiple self-media platforms, Kuaishou, and news platforms. It analysed the characteristics of self-media content, the mechanisms behind the formation of information silos, and the reinforcing role of algorithms. It also compared the silo behaviour exhibited by short video and news platforms. A measurement model for information silos was constructed based on Shannon's information entropy. Using collected user browsing data and recommendation information from news platforms, empirical calculations were performed to analyse overall user silo characteristics, differences in silo effects among users of varying activity levels, and the evolution of silo trends across different phases of the Kuaishou platform. Findings reveal that while information echo chambers exist across new media platforms, they do not form closed barriers. Higher information entropy values correlate with weaker echo chamber effects, and user activity levels exhibit a non-linear relationship with echo chamber intensity. Algorithm optimisation significantly mitigates echo chamber effects. This study's measurement model provides a basis for assessing echo chamber intensity, while its conclusions offer strategic support for mitigating information echo chambers.

Keywords

Online new media, information silos, measurement model, information entropy, recommendation algorithms.

1. Introduction

The flourishing development of internet platforms has reshaped the landscape of information dissemination. According to the Third National Time Use Survey Bulletin (2024), 92.9% of China's residents have established internet usage habits, with participants spending an average of 6 hours and 3 minutes online daily. With the number of internet users approaching 1.1 billion. Social media increasingly serves as the primary channel for users to receive content, where both the diverse voices of self-media and the professional output of public media are profoundly influenced by algorithmic recommendation mechanisms. This dependency has given rise to the phenomenon of information silos, where users' perspectives gradually narrow and their cognition becomes entrenched. Amidst the current era of information explosion and widespread personalised recommendations, this phenomenon not only impedes users' information acquisition and cognitive advancement but also poses potential challenges to the long-term development of online platforms. This paper's analysis of the factors influencing information silos and its quantitative modelling of impact intensity hold significant practical relevance.

Wang Fan et al. [1] examined the formation, impact, and mitigation of information silos within personalised recommendation systems. They proposed pathways including algorithmic optimisation and user guidance, constructing a multi-factor formation mechanism model. Their findings concluded that mitigating negative effects requires multi-stakeholder collaboration. Liu Qiang et al. [2] examined differences in the academic fields of "information silos" and "echo chambers" domestically and internationally. Collecting relevant literature data from CNKI and WOS between 2011 and 2020, they employed bibliometric methods and CiteSpace tools to construct a knowledge map model. Their findings revealed significant disparities in research priorities, clusters, and pathways between domestic and international studies. Zhu Hongtao et al. [3] examined the current state of information bubble research globally, proposing a content analysis methodology. By collecting relevant literature data from databases such as CNKI and WOS and employing a combined quantitative and qualitative approach, they constructed a multidimensional research review framework. Their findings indicated differences between domestic and international research, with the formation patterns and governance studies requiring further refinement.

Xu Xiang et al. [4] examined social network users' "cocooning tendency". Employing the BERT model to collect user post data from Sina Weibo, they constructed a predictive model through similarity calculations and regression methods. Their findings indicated that the degree of "cocooning" correlates with content similarity. Chen Si et al. [5] examined the adverse effects of information echo chambers on university students under intelligent algorithmic recommendation systems. They proposed approaches including enhanced regulation, strengthened information literacy education, and optimised algorithms, constructing a multi-stakeholder collaborative resolution model. Their findings indicate that multifaceted efforts are required to assist students in "breaking free from the cocoon". Zhou Xuan et al. [6] examined information silos within the scientific research domain under AI recommendation algorithms. They proposed strategies including algorithm transparency and optimised information distribution, constructing a multi-stakeholder collaborative response model. Their findings indicated that breaking out requires concerted efforts from policymakers, platforms, and researchers.

Wei Chunli et al. [7] addressed the challenge of measuring the intensity of information silos among social media users. They proposed the ICI-GCG method, constructing a user-generated content graph combined with multi-dimensional metrics. By collecting over 490,000 data points from 1,025 Sina Weibo users, their approach-utilising methods such as the BERT model-achieved greater measurement precision and enhanced interpretability. Xu Xiang et al. [8] examined information homogeneity within social media content production. Collecting 20.1174 million posts from 20,000 Sina Weibo users in 2018, they employed Word2vec, cosine similarity, and path analysis to construct a verification model. They concluded that higher "information bubble levels" among users correlate with greater alignment towards social information hubs, exhibiting homogenisation traits. Li Mao et al. [9] examined the deconstruction of human agency within digital-era "information bubbles," probing the roots of resulting pseudo-agency and anti-publicness. They proposed restoring human agency through optimising information production and dissemination, guiding opinion leaders, and cultivating netizens.

Liu Yuetan et al. [10] examined the information cocoon effect under algorithmic recommendation mechanisms using Toutiao as a case study. Analysing its impacts on individuals and society, they proposed regulatory approaches such as algorithm optimisation and emphasising "gatekeepers," constructing a multidimensional path to break free from the cocoon. Their conclusion emphasised the need for collaborative efforts between platforms and users to escape the cocoon's constraints. Cheng Yali et al. [11] examined methodological flaws in existing information bubble scales, proposing the concept of "algorithmic information bubble

susceptibility." Through two questionnaire surveys, they developed a 19-item scale comprising six indicators, which serves as an effective tool for micro-media effects research. Le Chengyi et al. [12] investigated the psychological motivations for short video users to break out of information silos under intelligent algorithmic recommendations. Through questionnaire surveys, they employed partial least squares structural equation modelling and multi-group analysis to identify factors influencing breakout intentions. Multiple factors were found to affect breakout intentions via emotional experiences, with content types exhibiting differentiated influences.

Xu Xiang et al. [13] examined information content convergence phenomena and effects within social media public information echo chambers. Employing text mining and statistical testing, they collected over 12.33 million posts spanning 726 days from Sina Weibo. After processing including Word2Vec vectorisation, they constructed a content convergence analysis model, revealing that information popularity is directly proportional to convergence and exhibits high synchronisation. Lu Xue et al. [14] investigated the "information echo chamber" effect among Weibo users and its influencing factors. Through questionnaire surveys collecting data from 619 users, they constructed a binary logistic regression model, confirming the widespread existence of "information echo chambers" on Weibo, influenced by factors such as gender. Yuan Rqing et al. [15] examined the organisational structure of fan communities and information dissemination characteristics within the Weibo ecosystem. Using Cai Chengyu's fanbase as a case study, they employed literature review, content analysis, and participant observation. Findings revealed clearly defined organisational divisions within fan communities, information dissemination guided by opinion leaders, and the presence of information echo chambers.

Hannák et al. [16] investigated the degree of personalisation in web search and its triggering factors. They proposed a method for measuring search result personalisation while controlling for noise, collecting data from Google, Bing, and DuckDuckGo. Using the Jaccard index and Kendall's tau coefficient for analysis, they constructed a multidimensional personalisation assessment model, concluding that both Google and Bing exhibit a degree of personalisation.

In summary, the information bubble effect has garnered considerable research attention, particularly with evidence of its existence identified in specific datasets. However, literature on quantifying the intensity of information bubbles remains relatively scarce. In this digital era of pervasive algorithmic recommendations, WeChat Video Accounts, leveraging its vast user base and diverse content ecosystem, has emerged as a pivotal platform for information dissemination. Yet, it also harbours risks of information bubbles at both individual and societal levels. This paper will collect data from WeChat Video Accounts to analyse the information bubble effect within short-video platforms and construct a measurement model for this phenomenon. On one hand, this research records user-level behavioural data-including likes, shares, and comments-to directly reflect individual information preferences and interaction tendencies. On the other, it collects dissemination metrics such as likes, shares, and comments on platform trending topics and public issues, clarifying the diffusion intensity and audience feedback of public information. Based on this data, a dual-dimensional data matrix of individual behaviour and platform dissemination is constructed. Employing content analysis and social network analysis methods, the formation logic and manifestations of personal information silos are examined, while the content characteristics and evolutionary patterns of public information silos are explored.

2. Analysis of User Information Acquisition Characteristics in the Context of Self-Media

The rapid development of internet platforms has fundamentally reshaped the underlying logic of information dissemination. As the core mechanism for content distribution on these

platforms, recommendation algorithms precisely match user preferences while simultaneously becoming a key driver of the information bubble effect. Under the continuous guidance of algorithms, users receive content that increasingly converges upon their existing spheres of interest, leading them into a predicament of narrowed perspectives and entrenched cognitive biases. This section first focuses on analysing the information echo chamber effect at the user level. By collecting user data from media platforms and comparing the differences in recommended content across various user groups, it delves into quantitative indicators of the echo chamber's impact intensity and methods for measuring its influence.

2.1. Analysis of Self-Media Content Characteristics

Presently, users' information acquisition pathways are undergoing profound, technology-driven transformation. The rapid advancement of internet technology and the widespread adoption of smart devices have provided unprecedented convenience and immediacy in information dissemination. Against this backdrop, users' media consumption preferences have fundamentally shifted away from traditional one-way channels like newspapers and television, towards interactive, fragmented, and personalised online media and self-media platforms such as Weibo, news apps, Douyin, and Xiaohongshu. This transformation is not an isolated phenomenon but is robustly supported by data. According to the Third National Time Use Survey Bulletin (2024), the average daily internet usage time among Chinese residents has reached six hours. This confirms that cyberspace has become the primary channel for the public to access information and has laid a solid user foundation for the subsequent flourishing of the self-media content ecosystem.

New media does not exist in a singular form but manifests as a diverse, layered landscape. Based on their functions and interaction models, they can be broadly categorised into three types: Firstly, the online extensions of official media, such as various news clients (People's Daily, Xinhua News Agency, etc.), whose core characteristics are high authority and rigorous content, albeit with relatively weak interactivity; Secondly, social interaction platforms, exemplified by WeChat and Weibo, emphasise peer-to-peer or point-to-multipoint communication among users, with information dissemination reliant on social relationship networks; Thirdly, short-video/live-streaming platforms, such as Douyin, Kuaishou, and Bilibili, whose content is characterised by strong visual impact, high entertainment value, and algorithmic driving forces, resulting in prolonged user engagement and heightened immersion. Table 1 summarises the key features of these three platform types.

Table 1. Analysis of Characteristics Across Three Categories of Online New Media

Platform Type	Representative Examples	Core Characteristics	Content Format
Online Versions of Official Media	People's Daily App, CCTV News	High authority, strong credibility, predominantly one-way dissemination	Text, images, and video news
Social Interaction Platforms	WeChat, Weibo	Characterised by social connections, strong interactivity, and rapid dissemination	Short text, images, links
Short Video/Live Streaming Platforms	Douyin, Bilibili, Xiaohongshu	Algorithm-driven recommendations, fragmented content, high immersion, strong interactivity, robust monetisation capabilities	Short videos, online livestreaming

The core characteristics of self-media content extend beyond its format, being deeply rooted in its unique economic drivers. Within this ecosystem, content creators, platforms, users, and advertisers/brands form a tightly knit community of shared interests. Creators produce

content to attract user engagement-likes, comments, shares-with user interaction metrics serving not only as direct indicators of content popularity but also as key gauges of its commercial value. Platforms employ sophisticated recommendation algorithms to precisely deliver high-engagement content to users with potential interest, thereby enhancing user retention and platform activity. Ultimately, surging popularity and traffic generate diversified revenue streams for creators, including platform traffic revenue sharing, advertising placements, and live-streaming sales commissions. Advertisers, as vital external participants, collaborate with creators either through the platform or directly for commercial placements. It is precisely this closed-loop mechanism, centred on "user preferences-engagement metrics-economic returns," that drives both content producers and platforms to strongly incentivise catering to-and even amplifying-users' existing interests. This dynamic lays the groundwork for the formation of "information silos."

2.2. Mechanisms Underlying the Information Bubble Effect

The information bubble phenomenon describes how users, influenced by their inherent preferences or algorithmic recommendation systems, consistently encounter content aligned with their existing views and interests. This prolonged exposure narrows their perspective and solidifies their cognitive framework, akin to being enclosed within a cocoon. This effect stems from the precision of algorithmic recommendations: platforms filter content based on user behaviour data, continuously reinforcing individual preferences while users selectively engage with content matching their inclinations. Technologically, personalised recommendation algorithms on platforms like Xiaohongshu and Weibo prioritise user retention and engagement duration. By continuously analysing click, like, and comment data, they precisely deliver homogeneous content, creating a feedback loop of "preference reinforcement – increased recommendations" that accelerates the solidification of the echo chamber. The existence of information silos poses significant hazards. For individual users, it restricts informational horizons, hinders comprehensive cognitive renewal and critical thinking development, potentially reinforcing biases and polarising viewpoints. For society, when the public becomes immersed in their respective information islands, the space for rational discussion on public issues shrinks, making social consensus difficult to achieve. This erodes the healthy ecosystem of the public sphere, posing potential risks to society's long-term stability and development.

The formation of information silos stems from multiple interrelated factors. Firstly, the explosive growth of the internet and the advent of the self-media era have generated an overwhelming supply of information and extreme fragmentation, providing fertile ground for users to selectively engage with content. Faced with information overload, users instinctively rely on personal interests for filtering, forming the psychological foundation for cocoon formation. Secondly, platform recommendation algorithms-designed primarily to maximise user engagement and commercial profit-constitute the most direct and potent technological engine constructing modern information cocoons. By continuously tracking and analysing user behaviour data-such as browsing, liking, and sharing-algorithms meticulously construct user profiles, subsequently favouring content highly aligned with historical preferences. This mechanism of catering to users' inclinations enhances experience while simultaneously weaving an invisible filter bubble that screens out diverse perspectives, trapping individuals within cognitive silos.

Thus, the formation of information silos is fundamentally a dynamic process where users' inherent preferences are continually identified, satisfied, and reinforced through algorithmic amplification. This is not an accidental phenomenon but a systemic product of contemporary information dissemination technology architecture and commercial models.

2.3. Recommendation Algorithms Reinforce the Cocoon Effect

The core purpose of recommendation algorithms across platforms is to precisely serve the commercial objective of enhancing user stickiness. To maximise user dwell time and engagement, algorithms are designed to continuously track user behaviour and prioritise content highly aligned with historical preferences. This stickiness-driven mechanism inevitably leads to a homogenisation of recommended content. Users become immersed within an algorithmically crafted "comfort zone" composed of similar content. While this delivers instant gratification, it simultaneously reinforces the information cocoon, progressively diminishing opportunities to encounter diverse perspectives and expand cognitive boundaries. Thus, the very purpose of these algorithms constitutes a key driver behind the increasingly pronounced cocoon effect.

Media platforms offer diverse, cross-disciplinary content provision. Each platform maintains multifaceted content offerings, presenting distinct thematic characteristics. Users can access multi-domain content across any platform. Table 2. summarises data covering 51 primary domain classifications across seven mainstream platforms. The table reveals: 1) Balanced content distribution, with the top 20 popular domains accounting for only 75.7% of total accounts, indicating active engagement in niche domains; 2) Cross-platform coverage: Popular domains average coverage across 6.2 platforms, demonstrating universal content provision.

Table 2. Cross-Platform Popular Category Statistics (Top 20)

Rank	Category	Total Accounts	Number of Platforms Covered	Platform Coverage
1	Food & Cooking	59	7	100%
2	Maternity and Parenting	52	7	100%
3	Technology & Digital	51	7	100%
4	Health and Wellness	45	7	100%
5	Gaming & Esports	45	6	86%
6	Home & Living	43	7	100%
7	Education & Examinations	43	7	100%
8	Film and Entertainment	42	6	86%
9	Romance & Relationships	37	6	86%
10	Career and Employment	37	6	86%
11	Photography	37	6	86%
12	Travel	35	7	100%
13	Beauty & Fashion	32	6	86%
14	Healthcare	32	6	86%
15	Automotive and Transport	31	5	71%
16	Lifestyle	31	6	86%
17	Legal System	28	5	71%
18	Shopping Recommendations	28	5	71%
19	Finance & Business	27	6	86%
20	History & Humanities	25	5	71%

Table 3 summarises the number of domains covered across seven major online platforms, including Weibo, alongside their primary domain names. The table reveals distinct focal areas for each platform: for instance, Bilibili emphasises gaming and technology, while Xiaohongshu concentrates on maternal and child-related content.

Table 3. Overview of Data Across Seven Major Platforms

Platform	Number of Domains	Primary Domains
Weibo	28	Beauty & Fashion, Technology Media, Finance & Investment
Zhihu	31	Technology & Digital, Education & Exams, Finance & Business
Xiaohongshu	18	Maternity & Parenting, Technology & Digital, Food & Cooking
Douyin	26	Beauty & Fashion, Film & Entertainment, Lifestyle & Home
Kuaishou	27	E-commerce Shopping, Health & Wellness, Food & Cooking
Bilibili	29	Gaming & Esports, Technology & Digital, Food & Cooking
Toutiao	34	Maternity & Parenting, Food & Cooking, Education & Exams

Table 4 presents statistics on 441 trending topics across 11 platforms, covering diverse fields including social welfare, entertainment, technology, and sports, reflecting the multifaceted nature of trending lists.

Table 4. Hot List Data Statistics

Platform	Chart Type	Number of Entries	Topics Covered
Weibo	Trending Topics	52	Entertainment, sports, etc.
Zhihu	Trending	50	Lifestyle, entertainment, etc.
Xiaohongshu	Trending	20	Entertainment, People's Livelihood, etc.
Douyin	Overall Chart	20	Entertainment, Public Welfare, etc.
Kuaishou	Real-time Hot List	50	Entertainment, Public Welfare, etc.
Bilibili	Sitewide Daily Chart	100	Entertainment, Technology, etc.
Baidu	Real-time Hot Topics	51	Sports, International, etc.
Today's Headlines	Trending Topics	50	Entertainment, International, etc.
The Paper	Trending	20	International, Technology, etc.
36Kr	24-Hour Hot List	8	Entertainment, Public Welfare, etc.
Minority	Popular Articles	20	Entertainment, Sports, etc.
Total	-	441	

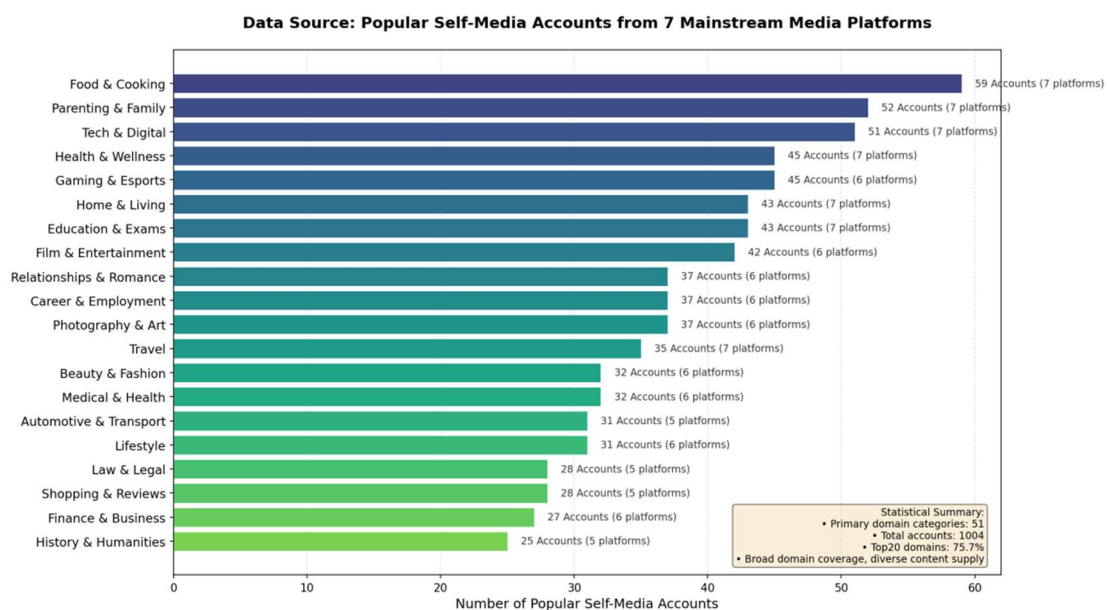
**Fig. 1** Cross-Platform Self-Media Content Domain Distribution (Top 20)

Fig. 1 illustrates the number of popular self-media accounts across primary domains. It demonstrates that content supply on media platforms is diverse and cross-domain, enabling users to access multi-domain content on any platform. This chart reveals that Food & Cooking, Maternity & Parenting, and Technology & Digital are the three domains with the widest cross-platform self-media presence, each boasting over 50 accounts across all seven platforms. The overall content ecosystem centres on lifestyle services while incorporating vertical tracks such as technology and entertainment, presenting a diversified supply pattern of "mass necessities + niche interests".

Fig. 2 Heatmap illustrating domain structure differences across platforms reveals that each platform offers diverse content supply, with distinct content characteristics. Users can access multi-domain content on any platform. Xiaohongshu concentrates on technology and digital, motherhood and parenting, and food and cooking domains, while Douyin focuses primarily on beauty and fashion. Each platform has developed unique content characteristics based on its domain layout, reflecting the stratified and differentiated development of new media platform content ecosystems.

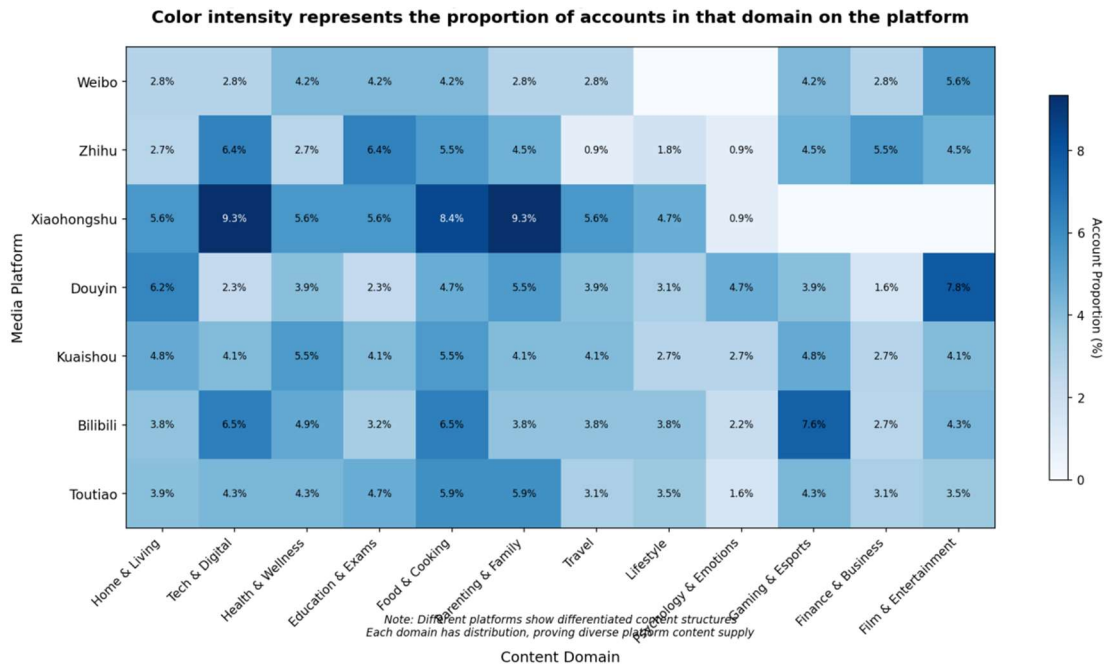


Fig. 2 Comparison of Popular Content Categories Across Platforms

Fig. 3 Based on real-time trending charts, this illustrates the distribution of popular content across domains, demonstrating that the content recommended by platforms is inherently diverse.

Fig. 3 illustrates the thematic distribution of trending content across platforms, clearly reflecting current patterns in online popularity rankings. Social issues and celebrity entertainment dominate the core of trending charts, accounting for the highest proportions and forming the primary content composition. These categories exhibit a marked disparity in popularity compared to fourth-ranked and subsequent themes. Conversely, education, culture, and international affairs feature relatively low representation in trending content, demonstrating significantly weaker engagement. Overall, while platform trending charts cover multiple domains, their popularity distribution exhibits marked imbalance. Lifestyle and entertainment topics, which command high public attention, dominate the charts, whereas professional and knowledge-based domains receive comparatively limited exposure.

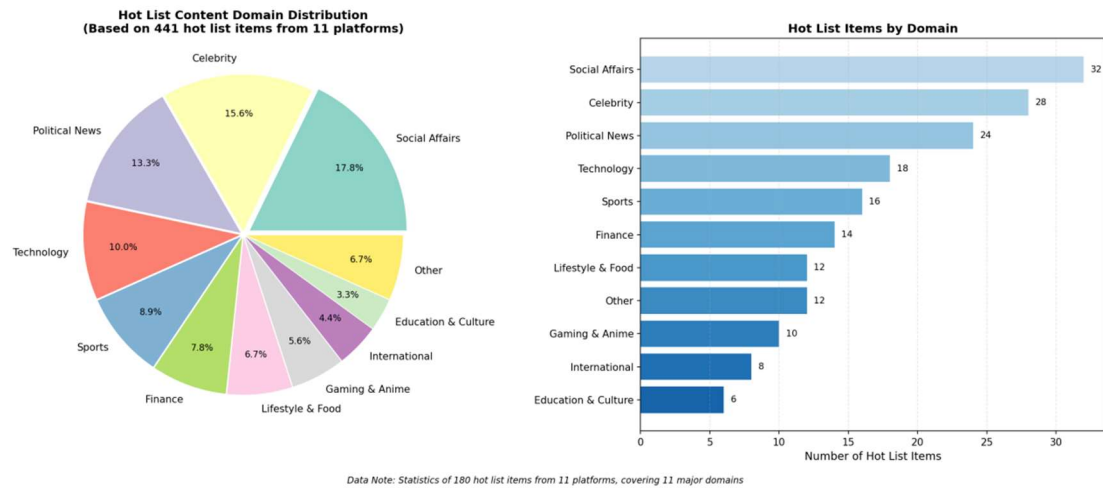


Fig. 3 Analysis of Content Domain Diversity in Trending Charts Across Platforms

3. Analysis of Content Characteristics on Short Video and News Platforms

In an era where algorithmic recommendation technologies deeply permeate information consumption scenarios, platforms like News and Kuaishou-as highly influential domestic information and social media platforms-are profoundly shaping public cognitive landscapes through their information distribution mechanisms. This chapter will systematically examine the role of recommendation algorithms in forming and reinforcing information silos, based on user behaviour data from these two platforms. We will conduct empirical analysis across multiple dimensions-including historical user preferences, content similarity, and breadth of diverse information exposure-to reveal how personalised recommendations immerse users in homogenised information loops. This ultimately leads to profound societal impacts such as group opinion polarisation and diminished engagement with public issues.

3.1. Kuaishou Data Analysis

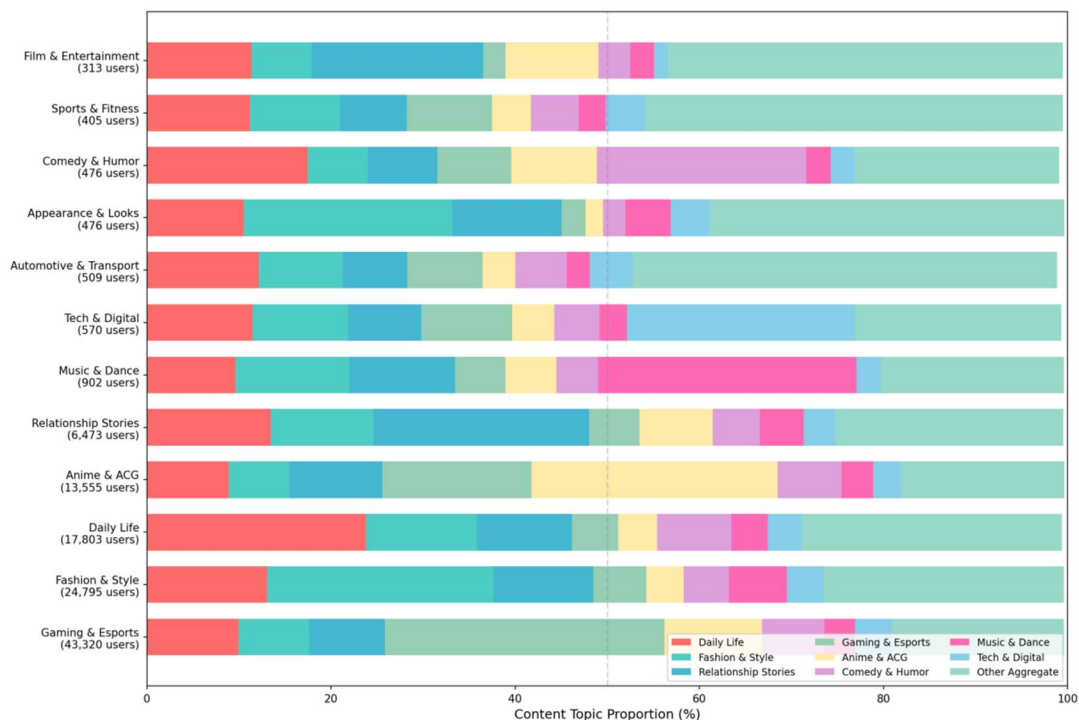


Fig. 4 Thematic Distribution of Content Consumption Among Heavy Interest Users Across Domains

This section utilises publicly available Kuaishou platform data to demonstrate that the information bubble effect is indeed present, though such bubbles are not "completely sealed off"; cross-domain content consumption remains prevalent. The data reveals that heavy users across various domains typically consume content within their primary domain at a rate between 23% and 30%, significantly higher than consumption within any single domain.

Fig. 4 illustrates the thematic composition of content consumed by heavy users across 12 major interest domains. Analysis reveals that content consumption across all domains exhibits diversification, with domain-specific content holding the highest share yet typically not exceeding 30%. Clear domain correlations exist (e.g., gaming and anime; fashion and aesthetics).

Fig. 5 illustrates the proportion of heavy users consuming "content within their own domain" across various sectors. Data analysis indicates that the share of domain-specific consumption ranges between 23% and 30% across all sectors, with none exceeding 50%. This demonstrates that information silos are not entirely closed systems. Heavy users in the music and dance sector exhibit the highest proportion of domain-specific consumption at 28.1%.

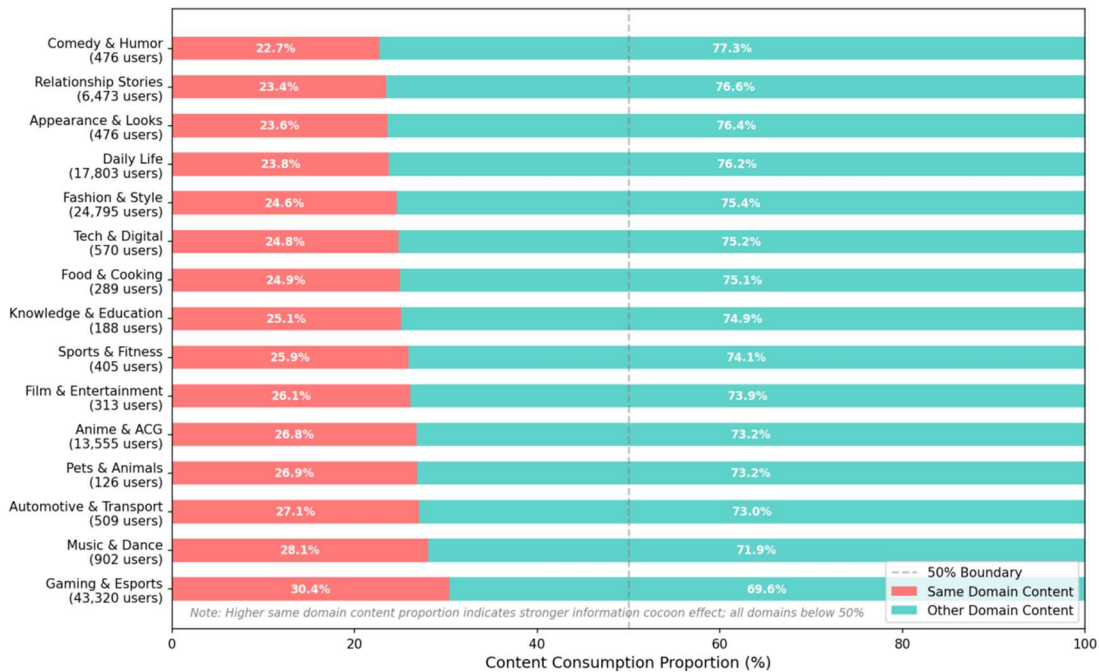


Fig. 5 Proportion of "Content Within the Same Domain" Consumed by Heavy Users Across Domains

Data source: Kuaishou OpenOneRec public dataset. Dataset URL:
<https://huggingface.co/datasets/OpenOneRec/OpenOneRec-RecIF>.

3.2. Analysis of News Platform Data

This section analyses data collected on heavy interest users within the news domain. Findings reveal that recommendation algorithms significantly exacerbate the information bubble effect by continuously reinforcing user interests, trapping users' click behaviour within their primary interest domains. Heavy interest users' clicks are highly concentrated within their main domain, with an average primary domain concentration of 21.1% across 20 analysed domains.

Fig. 6 illustrates the click behaviour characteristics of heavy interest users across different domains. Red bars represent the user's primary interest domain, while blue bars denote clicks on other domains. The proportion of clicks within the primary domain for heavy interest users in each domain did not exceed 50%, generally ranging between 20% and 33%. This indicates that users do not consume content from a single domain exclusively but widely click on content

from other domains such as film and television, animals, and society, suggesting that interest boundaries are not closed.

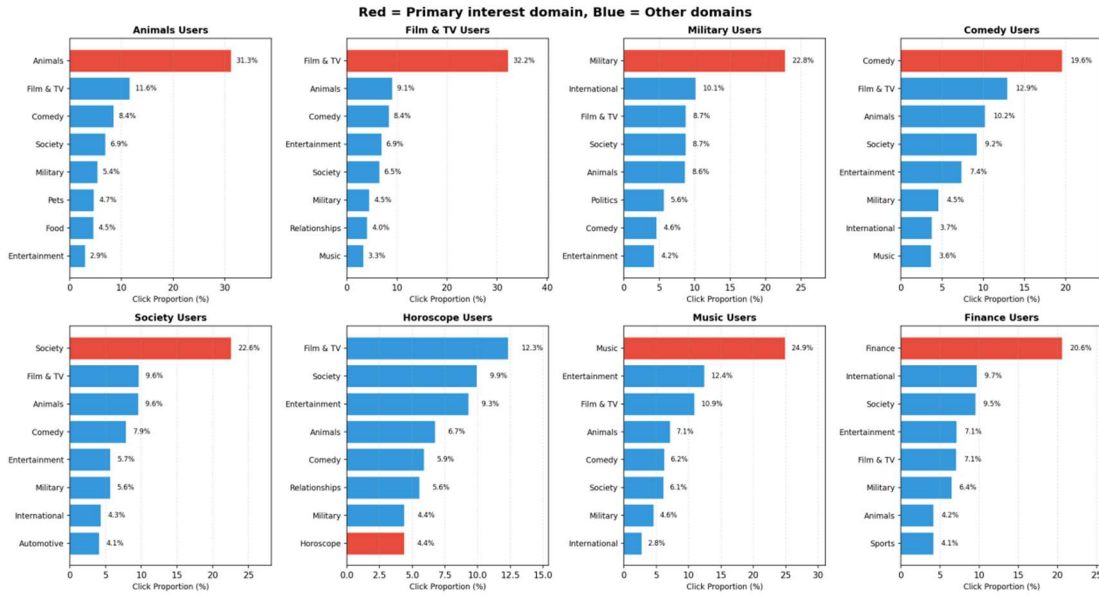


Fig. 6 Click Topic Distribution of Heavy Interest Users by Domain

4. Construction and Analysis of the Information Bubble Effect Measurement Model

Within the information bubble effect measurement model, entropy serves as the core metric for assessing the diversity of users' information consumption. This study adopts Shannon's information entropy as the primary metric for assessing information bubble intensity. This choice aligns with entropy's fundamental role in quantifying the uniformity of distribution, precisely matching the study's definition of an information bubble—namely, whether users' content consumption is concentrated within a limited number of domains. It accurately corresponds to the consumption proportion characteristics across different content domains observed on news and short-video platforms, demonstrating high suitability for the research context. Furthermore, the calculation of information entropy in this study directly translates into the proportion of consumption across various content domains for each user. This computational logic is clear and operational, making it suitable for empirical analysis. A higher entropy value indicates greater diversity in the information domains users engage with and more dispersed content sources, signifying a weaker information bubble. Conversely, a lower entropy value suggests users' information consumption is highly concentrated within a few domains, with greater homogeneity in content selection, indicating a stronger information bubble. By quantifying the diversity of information consumption, this model provides an operational assessment basis for determining whether users are trapped within an information bubble.

4.1. Construction of an Information Bubble Effect Measurement Model Based on Information Entropy

This paper primarily employs Shannon's information entropy to measure the strength of information silos.

Defining core symbols: Let U denote an individual user, $X=\{x_1, x_2, \dots, x_n\}$ represent the set of information content encountered by that user, $p(x_i)$ denote the proportion of the i -th content

($\sum_{i=1}^n p(x_i) = 1$, $0 \leq p(x_i) \leq 1$), and $H(U)$ denote the information bubble metric. The mathematical model is: $H(U) = -\sum_{i=1}^n [p(x_i) \times \log_2 p(x_i)]$, where $H(U) \geq 0$.

Information entropy exhibits a negative correlation with the information echo chamber: the larger $H(U)$, the more uniformly distributed $p(x_i)$ becomes, indicating greater diversity in user information content and a weaker echo chamber effect; the smaller $H(U)$, the more concentrated the information content, resulting in a stronger echo chamber effect; when $H(U)=0$, the echo chamber effect is at its strongest.

4.2. Exemplary Calculations of the Information Cocoon Measurement Model

4.2.1. Data Overview

Data source: Aggregated statistical data on user content topic distribution from mainstream domestic news platforms

Fig. 7 indicates that the mean entropy of user topic distribution is 1.56, with a median of 1.74. This suggests that overall user interest distribution exhibits a degree of concentration, indicating that the information bubble effect is not extremely pronounced but has not been entirely mitigated. The mean number of effective topics is approximately 6.3, meaning users' attention is "equivalent to" being evenly distributed across roughly six topics, reflecting a degree of diversity. However, the average proportion of primary topics is 42%, indicating that nearly half of users' content consumption remains concentrated on a single topic, with interest boundaries still exhibiting a pronounced core-focused characteristic.

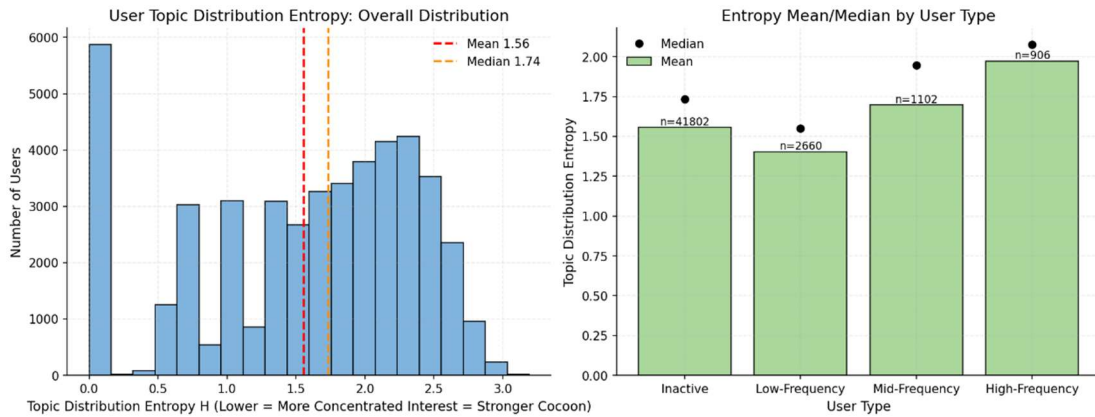


Fig. 7 Entropy distribution of user topic distribution

4.2.2. Information Bubble Characteristics Across User Activity Levels

Comparing information bubble metrics across user types reveals that low-frequency users exhibit the lowest theme distribution entropy (1.40) and effective theme count (5.73), alongside the highest proportion of primary themes (46.4%), indicating a relatively stronger information bubble effect. Conversely, high-frequency users exhibited the highest topic distribution entropy (1.97) and number of effective topics (8.28), alongside the lowest proportion of primary topics (36.5%). However, their self-circulation probability (25.2%) was significantly higher than other user types, indicating that while their content consumption is more diverse, they are also more prone to forming closed loops within their own areas of interest. Overall, user activity levels and information bubble intensity do not exhibit a straightforward linear relationship. Low-frequency users are more prone to concentrated consumption within a single topic, while high-frequency users, despite engaging in diverse consumption, demonstrate stronger self-circulation characteristics within their domains.

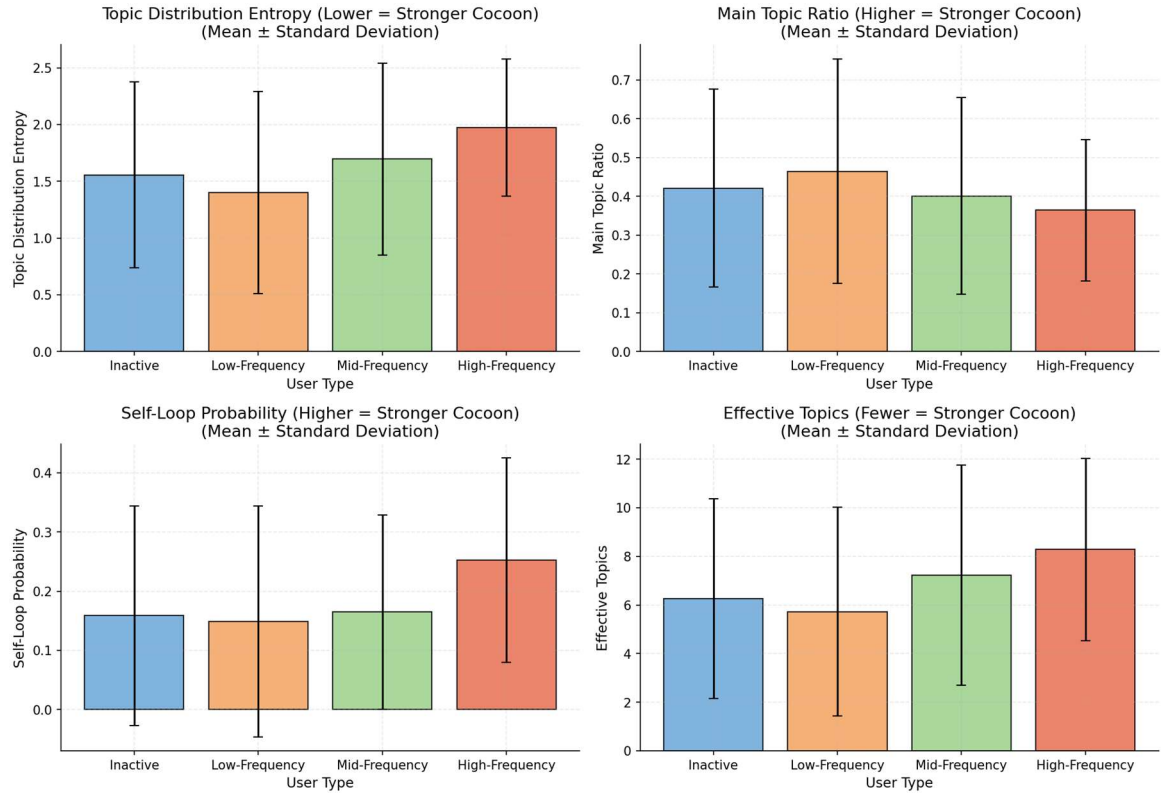


Fig. 8 Composite Metrics by User Type

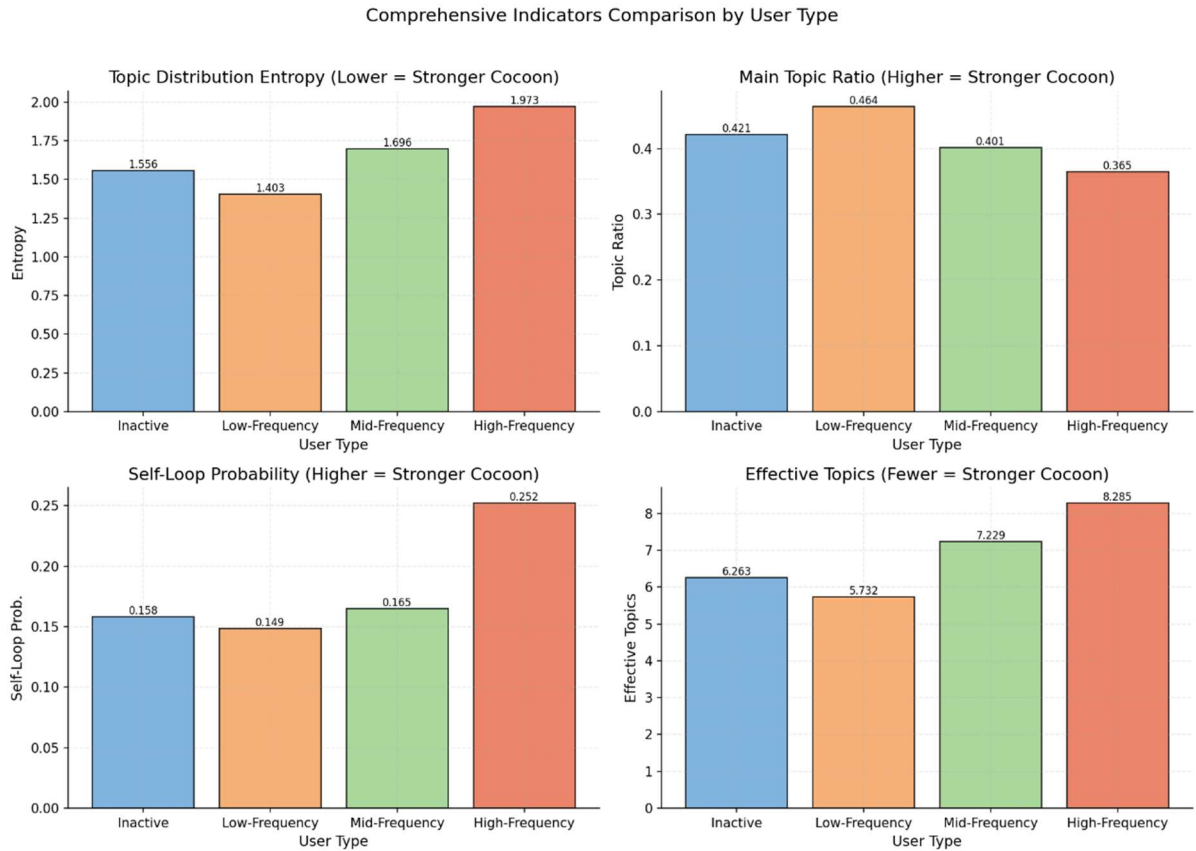


Fig. 9 Comparison of Composite Metrics

4.3. Case Study Analysis of Information Bubble Intensity Trends

Table 5. Quantitative Differences in Information Silos Across Different Time Periods on Kuaishou

Analysis Phase	Quantitative Performance of Information Entropy (H(U))	Proportion of Core Interest Content	Core Quantitative Metrics	Group Variation Characteristics
Before March 2025	Mean 1.02 ± 0.18 bits; Strong echo chambers accounted for 48.7%, weak echo chambers only 1.0%; Coefficient of variation 17.6%	Mean $72.3 \pm 5.8\%$; 19.2% of users exposed to only 1-2 content categories	Echo chamber intensity: 82.5%; $R^2 = 0.89$ for over-personalisation fit	Differential coefficient: 12.3%; H(U) difference between younger/older users: 0.09 bits ($P > 0.05$)
2025.3-5	Increased from 1.02 to 1.93 bits (89.2% rise); coefficient of variation 13.0%	Decreased from 72.3% to 54.8% (17.5 percentage point reduction)	Eggshell intensity reduced to 47.2%; proportion of strong eggshells reduced to 12.1%	Differential coefficient 28.7%; H(U) difference between cocoon-breaking users and non-users 0.76 bits ($P < 0.05$)
Post-2025.5 to 2026.2	Mean 2.17 ± 0.22 bits; weak cocoon rooms accounted for 60.0%; CNY cycle increased to 2.3–2.5 bits	Mean $45.6 \pm 4.9\%$; balanced content distribution	Echo chamber intensity: 31.8%; cross-domain content increased by 220% compared to previous period	Coefficient of variation: 20.5%; Gap between older/lower-tier users and younger users narrowed to 0.23 bits

Combining the aforementioned quantitative data yields the core conclusion: Kuaishou's information silo effect exhibits a marked weakening trend alongside algorithmic optimisation. The disparity in data across different phases fundamentally stems from algorithmic strategy iteration and enhanced user autonomy. Prior to algorithmic refinement, users experiencing strong echo chambers constituted nearly half the user base, with minimal differences in echo chamber effects between groups—primarily driven by a singular precision-based recommendation algorithm. During the transitional phase, information entropy rapidly increased, significantly weakening the echo chamber effect. Differences in user behaviour and group characteristics became pronounced, with the 'break-the-cocoon' function beginning to exert its influence. During the stabilisation phase, information entropy converged towards a reasonable range, with weak echo chambers becoming predominant and group disparities tending towards equilibrium. The alignment of algorithmic benevolence with content ecosystem refinement effectively mitigated the echo chamber effect, broadly consistent with the quantified negative correlation between information entropy and echo chamber effects.

5. Conclusion

This study primarily examines the causes, quantification methods, and evolution patterns of the information cocoon effect within self-media environments. Empirical analysis was

conducted using user behaviour data from Kuaishou and news platforms, alongside monitoring results across different time periods. Current core findings indicate that while information cocoons are prevalent on short-video and news platforms, they have not formed entirely closed cognitive barriers. Users still encounter content from diverse domains. The information entropy measurement model, grounded in Shannon's information theory, effectively quantifies the intensity of information silos, with higher entropy values indicating weaker silo effects. The core solution lies in algorithmic optimisation: altering traditional recommendation models to grant users greater control over content preferences, while increasing support for high-quality, diverse content to mitigate the impact of information silos. This study also presents certain limitations, such as insufficient sample coverage, superficial analysis of external factors like individual user characteristics and social environments, and room for further refinement in the constructed measurement model. Future research could expand sample coverage, delve deeper into the impact of multi-factor interactions on information silos, continuously improve the measurement model, and explore the practical application of relevant technologies in silo governance.

References

- [1] Wang Fan. Formation Mechanisms, Negative Impacts, and Mitigation Pathways of Information Echo Chambers in Personalised Recommendation Contexts [J]. *E-Commerce Review*, 2025, 14(4):1591-1598.
- [2] Liu Qiang, Wan Bowen, Yu Han. Information and the Cocoon: A Comparative Analysis of Domestic and International Academic Discourses on the "Information Cocoon" [J]. *Journal of Education Media Research*, 2024(2):81-89.
- [3] Zhu Hongtao, Li Shuxi. A Review of Research on Information Echo Chambers [J]. *Library and Information Work*, 2021, 65(18):9. DOI:10.13266/j.issn.0252-3116.2021.18.015.
- [4] Xu Xiang, Yu Junjun. User "Cocoon Homogeneity" in Social Network Content: A Sample Study of Sina Weibo Based on the BERT Model [J]. *Journal of Beijing Institute of Technology: Social Sciences Edition*, 2023, 25(4):182-191. DOI:10.15918/j.jbitss1009-3370.2022.3989.
- [5] Si Chen. Negative Impacts of Information Cocoon Phenomena on University Students in the Context of Intelligent Algorithm Recommendations [J]. *Advances in Social Sciences*, 2023, 12(08):4261-4267. DOI:10.12677/ass.2023.128581.
- [6] Zhou Xuan. Artificial Intelligence Recommendation Algorithms and the Information Cocoon Phenomenon: Challenges and Countermeasures in Scientific Information Acquisition [J]. *Modern Social Science Research*, 2025; 5: (2): 119-122.
- [7] Wei Chunli, Xu Jianmin, Zhang Guofang. A Study on Measuring the Strength of Social Media Users' Information Echo Chambers Based on Generative Content Graph Analysis [J]. *Journal of Information Science*, 2025(10).
- [8] Xu Xiang, Ao Ziqi, Shi Jingyuan, et al. Divergent Paths, Convergent Outcomes: Convergence of Information Echo Chambers in Social Media User Content Production - An Empirical Analysis Based on Sina Weibo [J]. *Journal of Xi'an Jiaotong University (Social Sciences)*, 2022, 42(3). DOI:10.15896/j.xjtuskxb.202203015.
- [9] Li Mao, Han Pugeng. Deconstruction and Reconstruction of Subjectivity Under the Constraints of the "Information Cocoon" in the Digital Age [J]. *Jiangsu Social Sciences*, 2024(3).
- [10] Liu Yuetan, Wang Xuemin. Reflection and Regulation of the Information Cocoon Effect under Algorithmic Push Mechanisms [J]. *Media*, 2024(13).
- [11] Cheng Yali, Huang Youjia. Development and Validation of an Algorithmic Information Cocoon Susceptibility Scale [J]. *Journalism World*, 2024(3):50-61.
- [12] Le Chengyi, Zhang Jinping, Wang Zixin, Chen Tao. Escape or Entrapment? Psychological Motivations for Short Video Users to Break Through Information Echo Chambers Under Intelligent Algorithm Recommendations [J]. *Modern Intelligence*, 2025, 45(8):43-56.

- [13] Xu Xiang. Public Information Echo Chambers: Convergence Phenomena and Effects in Social Media Content – A Text Mining Study Based on Sina Weibo [J]. Journal of Information Science, 2023, 42(1):113-123.
- [14] Lu Xue, Sun Lei, Zheng Yimin. Research on the "Information Cocoon" Effect and Influencing Factors Among Weibo Users [J]. China Market, 2022(16).
- [15] Yuan Ruching. A Brief Analysis of Fan Group Organisational Structures and Information Dissemination Characteristics in the Weibo Field: Taking Cai Chengyu Fans as an Example [J]. Shanghai Radio and Television Research, 2021(3):6.
- [16] Hannák A, Sapieżyński P, Molavi Kakhki A, et al. Measuring Personalisation of Web Search [C]//Proceedings of the 22nd International Conference on World Wide Web. Rio de Janeiro: ACM, 2013:527-538.