

Algorithmic Gatekeeping And Consensus Fragmentation: How Platform Curation Shapes Public Discourse On Social Media News

Kaixiang Yang^{1*}, Manual Selvaraj Bexci²

¹School of Social Sciences, Arts and Humanities, Lincoln University College, Petaling Jaya, No.47301, Selangor, Malaysia

^{1*}kaixiang@lincoln.edu.my

²School of Social Sciences, Arts and Humanities, Lincoln University College, Petaling Jaya, No.47301, Selangor, Malaysia.

Abstract: In the era of algorithm-driven platform communication, social media platforms have become the core channel for the public to access news and information, and the algorithmic gate-keeping mechanism embedded in them is reshaping the logic of public discourse production and the basis of consensus building in an unprecedented way. Taking 'algorithmic public sphere' as the core analytical framework, this study adopts a mixed-methods research design, integrating systematic literature synthesis, computational text analysis, and critical discourse analysis, to systematically examine the discourse corpus (N≈1.2 million items) on three major platforms, namely Twitter/X, Facebook, and YouTube, on three types of issues, namely, political elections, public health, and climate change, between 2020 and 2024. 1.2 million) is systematically examined with the aim of revealing how platform curation mechanisms lead to the fragmentation of public consensus. The study finds that: first, algorithmic high push content is significantly higher in affective intensity than low push content (Cohen's $d = 0.47-0.61$), and conflicting discursive frames systematically dominate algorithmically prioritised distribution of content (43.7% on average), revealing the structural exclusion of engagement-oriented algorithmic logics from rational public discourse; second, the discourse dissemination networks of the three platforms show a highly modular clustering structure ($Q = 0.63-0.71$), with the proportion of cross-community bridging connections ranging from 8.3% to 11.7%, indicating that the algorithmic gate-keeping mechanism objectively strengthens the tendency of the discourse community's closure; third, the 'factual definitions' of the same public issue in different discourse communities have a 4- to 6-fold difference in frame frequency, and the fragmentation of consensus has already occurred. Thirdly, there is a four to six times higher frequency of disagreement between different discourse communities on the 'factual definition' of the same public issue, and the fragmentation of consensus has gone beyond the level of emotional polarisation to the epistemological dimension. At the theoretical level, on the basis of critically inheriting Habermas's theory of the public sphere, this study defines the algorithmic gatekeeping mechanism as the 'level of platform technology design' not covered by traditional gatekeeping theories, and preliminarily constructs the analytical dimensions of the 'differentiated platform discourse theory'. At the policy level, the findings of the study are of direct reference value to the regulation of algorithmic transparency, the reconstruction of journalistic professionalism, and the cultivation of digital citizens' algorithmic literacy under the framework of the EU's Digital Services Act

Keywords: algorithmic gatekeeping; platform curation; public discourse; consensus fragmentation; filter bubbles; echo walls; digital public sphere

Introduction

In the era of digital communication, the mode of production, distribution and consumption of news and information is undergoing a profound structural change. Social media platforms have replaced traditional media organisations as the primary channel for the public to obtain news information. According to the Pew Research Centre, by 2023, more than half of adults worldwide will.

use social media as one of their daily news sources (Shearer & Mitchell, 2021). However, unlike traditional media that rely on human editors for content selection, the current news dissemination ecosystem is largely constrained by automated algorithmic systems - which are based on user behavioural data and rely on machine learning techniques to filter, sort and push massive information streams in real time (Napoli, 2019). The intervention of algorithmic mechanisms has profoundly reshaped the boundaries and attributes of public discourse spaces.

The rise of the concept of 'algorithmic gatekeeping' (algorithmic gatekeeping) marks a significant shift in the understanding of information flow control in the field of communication. While traditional gatekeeping theory concentrates the power of information screening in the hands of professional practitioners such as journalists and editors (Shoemaker & Vos, 2009), algorithmic gatekeeping shifts this power to automated systems designed by platform owners and their engineers (Gillespie, 2014). Facebook, Twitter/X, TikTok, and other mainstream platforms have reinforced the logic of engagement-oriented content distribution by means of algorithmic recommendation mechanisms, so that emotionally compelling, hard-hitting news content receives a higher priority for dissemination (Fisher, 2022). The deeper consequence of this mechanism is the homogenisation and polarisation of the information ecology to which the public is exposed, resulting in a systematic erosion of the pluralistic public discourse necessary for democratic consultation.

"Consensus fragmentation is the central concern of this study. In the ideal public sphere envisioned by Habermas (1989), citizens form public opinions through open and rational dialogue, thereby constructing social consensus. However, algorithm-driven platform environments, through the mechanisms of 'filter bubble' and 'echo chamber', close users in an information cocoon (Pariser, 2011; Sunstein, 2017), weakening encounters and encounters between different viewpoints, resulting in an increasingly barren ground for the formation of social consensus. In recent years, several large-scale empirical studies on platforms such as Facebook and Twitter have demonstrated a significant correlation between algorithmic content recommendation and the polarisation of users' political opinions (Bail et al., 2018; Gonzalez-Bailon et al., 2023), although the controversy over the causal mechanism has yet to subside.

Although a considerable body of literature has explored the relationship between algorithms and news dissemination from a technological, political, or psychological perspective, several important gaps remain in existing research. First, most studies focus on the influence of algorithms on political beliefs or voting behaviour, while there is a lack of systematic analysis of the constructive role of public discourse itself. Second, platform algorithms are not neutral technological tools, and their design logic is embedded with specific commercial interests and ideological presuppositions, but existing studies have paid insufficient attention to the political economy dimension of algorithms (Bucher, 2018; Van Dijck et al., 2018). Third, in terms of theoretical framework, how to organically integrate the classical theories of traditional communication with the new features of algorithmic communication is still a theoretical project that needs to be deepened.

In view of this, this paper aims to explore the following core research questions through a systematic literature review and critical analysis framework: first, how do algorithmic gatekeeping mechanisms intervene and reshape the process of public discourse production in social media news communication? Second, under what mechanism does platform curation lead to the fragmentation of public consensus? Third, what does the current pattern of discourse fragmentation mean for democratic consultation and social cohesion?

The theoretical contribution of this study is at two levels: firstly, it attempts to construct an analytical framework of the 'algorithmic public sphere' based on the existing gatekeeping and public

sphere theories; secondly, it provides a more comprehensive theoretical explanation of the institutional causes of consensus fragmentation by integrating the research perspectives of computational communication and critical communication. This paper is structured as follows. This paper is structured as follows: Section 2 systematically reviews the existing research in the four core areas; Section 3 introduces the research design and methodology; Section 4 presents the main findings; Section 5 discusses the theoretical and practical implications; and Section 6 draws conclusions and points out the direction for future research.

Literature Review

2.1 Evolution of Gatekeeping Theory: From Manual Editing to Algorithmic Dominance

Gatekeeping theory is one of the most influential classic theories in communication, and its theoretical roots can be traced back to the 'food channel' metaphor proposed by social psychologist Kurt Lewin in 1943 (Lewin, 1943). White (1950) introduced this framework into journalism research, using the case study of 'Mr. Gates' to reveal the role of personal judgement and subjective preference in editorial decision-making, and established the empirical tradition of gatekeeping journalism. Shoemaker & Vos (2009) refined the gatekeeping process into a multi-level system model, covering five levels: individual, routine process, organisation, social institution and social system.

However, with the rise of the Internet and social media, the mediated environment on which traditional gatekeeping theory is founded has fundamentally changed. Singer (2014) points out that the spread of digital technology has made it possible for any individual with Internet access to become a publisher of news content, and the physical significance of 'gatekeeping' has been greatly dissolved at the technological level. At the same time, platform algorithms are gradually filling the void left by the retirement of traditional editorial authority (Napoli, 2014). Gillespie (2014) pioneered the proposition of the relevance of algorithms, suggesting that algorithms are not only technological tools, but also political "gatekeepers of visibility".

There are three important structural differences between algorithmic and traditional human gatekeepers: differences in scale and speed, differences in transparency (Pasquale, 2015) and differences in personalisation (Zuiderveen Borgesius et al., 2016). Bucher (2018) further suggests that algorithms regulate user behaviour by predicting it and creating "threats of engagement", creating a new type of power technology. Starting from a "platform society" framework, Van Dijck, Poell & De Waal (2018) argue that platform algorithms infiltrate and reorganise domains with public attributes in the logic of commercial data, creating a deep social change of "platformization".

2.2 Filter Bubbles and Echo Walls: Mechanisms, Evidence and Controversies

In his book of the same name, Pariser (2011) pioneered a systematic description of the concept of "filter bubbles", suggesting that personalisation algorithms continually track users' behaviour, blocking them in a cocoon of information that the system predicts to be "of interest" to them. Sunstein (2017) focuses on the phenomenon of the "echo chamber" in his "Republic.com" series of books, highlighting the tendency of users, driven by both algorithmic recommendations and active choices, to expose themselves to information that conforms to their pre-existing views, thus increasing polarisation at the political level.

The above theory is partially supported by several widely influential empirical studies. Bakshy et al. (Bakshy, Messing & Adamic, 2015), based on data analyses of over 10 million Facebook users, found that algorithmic sorting had a more significant role in limiting users' exposure to heterogeneous political content. Through a randomised experiment, Bail et al. (Bail et al., 2018) found that exposure to

information across political boundaries not only failed to reduce polarisation, but further exacerbated the reinforcement of ideological positions among groups of conservatives. A Facebook experiment by Gonzalez-Bailon et al. (Gonzalez-Bailon et al., 2023), published in *Nature*, found that algorithmic push substitution for chronological ordering reduced users' exposure to politically extreme content, but had a very limited effect on changing the users' political views themselves.

However, the filter bubble and echo wall theories have also been questioned and refuted by a large body of empirical research. Flaxman et al.'s (Flaxman, Goel & Rao, 2016) analysis of web browsing data found that social media actually increased users' exposure to heterogeneous content. Prior's (2013) study, on the other hand, suggests that users' active choices have a non-negligible dynamic in shaping the information environment. Nechushtai et al. (Nechushtai & Lewis, 2019) point out that existing studies are fundamentally divergent in terms of conceptual definitions, measurement approaches and causal inferences. Möller et al. (Möller et al., 2018) argue that filter bubbles are moderated by multiple factors such as platform design, content type, and user characteristics.

2.3 Public domain theory and discourse production in the algorithmic era

Habermas' (Habermas, 1989) theory of the public sphere (public sphere) provides the central normative framework for this study. In his latest work, *The New Structural Transformation of the Public Sphere*, Habermas (2022) explicitly responds to the challenges posed by digital media to the public sphere, viewing the fragmented communication ecology of algorithmic platforms as a systemic threat to rational public discourse. Fraser (1990) points out that the bourgeois public sphere as conceived by Habermas fundamentally excludes the disadvantaged, and that algorithmic systems are similarly systematically biased in the context of digital communication (Noble, 2018). Papacharissi (2002, 2015), on the other hand, notes from an optimistic perspective that the Internet provides a new living space for the "counterpublic" (counterpublic).

In the context of algorithmic platforms, the production of news discourse follows a unique logic. McCombs & Shaw's (1972) agenda-setting theory reveals the mechanism by which traditional media influence public perceptions through reporting on issues; in the age of algorithms, platform algorithms actually play the role of a 'meta-agenda setter' by dynamically adjusting the distribution of content traffic (Meraz, 2011). Entman (1993) defines framing as the selective accentuation of certain aspects of reality; algorithms prioritise the delivery of high-engagement content as a systematic reward for emotional, conflicting frames (Brady et al., 2017). Content creators' "algorithmic literacy" is increasingly an important threshold for discourse acquisition (Bucher, 2018).

2.4 Consensus fragmentation: theoretical explanations and practical consequences

'Consensus fragmentation' is the core analytical concept proposed in this study, aiming to describe the overall trend of public discourse in the algorithmic platform environment towards the fragmentation of pluralistic and isolated meaning systems. While political polarisation usually refers to the tendency of political views to move in opposite directions at both ends of the ideological spectrum (Iyengar et al., 2019), consensus fragmentation is more concerned with the disintegration of shared understanding and shared reality among discourse subjects (Eady et al., 2019). 2019).

The mechanisms by which algorithmic gatekeeping contributes to the fragmentation of consensus can be summarised at three levels: first, the fragmentation of the information environment (Lewandowsky et al., 2017); second, the intensification of affective mobilisation (Brady et al., 2017; Vosoughi, Roy & Aral, 2018); and, third, the closure of the discourse community (Cinelli et al., 2021). The real-life consequences of the above mechanisms are evident in a number of public issue areas such

as climate change, public health (Roozenbeek et al., 2020), and political elections (Iyengar et al., 2019).

2.5 Research Gaps and Theoretical Orientation of this Paper

Synthesising the above literature review, there are three theoretical gaps in existing research that need to be urgently filled: first, focusing on effect rather than process (Nechushtai, 2018); second, insufficient theoretical integration (Freelon, 2015); and third, ignoring platform differences (Taneja et al., 2018). This study takes 'algorithmic public sphere' as the core analytical framework, integrates Habermas's normative theory, the mechanism perspective of gatekeeping theory and the political economy dimension of platform research, and systematically explains how algorithmic gatekeeping leads to consensus fragmentation.

Research Methodology

3.1 General Framework of Research Design

This study adopts a mixed methods research design, integrating systematic literature synthesis, critical discourse analysis (CDA) and computational text analysis (CTA). computational text analysis (Creswell & Plano Clark, 2018). The synergistic use of the triple methodological pathway aims to construct a methodological triangulation of the research findings to enhance the confidence and validity of the analytical conclusions (Flick, 2018). The philosophical stance of this study follows the epistemological orientation of critical realism (Bhaskar, 1998), and the core task of the study is to reveal the deep causal structure of 'why it happened' (Archer et al., 1998).

3.2 Data sources and sample selection

3.2.1 Literature Corpus Construction

In this study, a systematic corpus of literature covering research on algorithmic gatekeeping and public discourse was first constructed. The search databases include Web of Science, Scopus, Google Scholar, and Communication Abstracts, and after preliminary screening and full-text reading, a total of 87 pieces of intensively read literature are finally included, including 74 SSCI-indexed journal articles and 13 monographs and book chapters. Literature quality assessment prioritised the inclusion of research results published in high-impact journals such as *New Media & Society*, *Journal of Communication*, *Political Communication*, *Information, Communication & Society*.

3.2.2 Social Media News Discourse Dataset

In this study, news-related content in the comment sections of Twitter/X, Facebook and YouTube was selected as the object of discourse analysis, focusing on three types of issues between 2020 and 2024: political election issues, public health issues (the new crown epidemic) and climate change issues. The final analysed corpus contains approximately 1.2 million posts/comments extracted from the original dataset by stratified random sampling method (stratified random sampling), stratified by platform (Twitter: 40%; Facebook: 35%; YouTube: 25%) and issue type (Krippendorff, 2018).

3.3 Analysis methods

3.3.1 Computational text analysis

This study employs computational text analysis to systematically describe the discourse corpus, and the specific analysis procedure includes four segments: first, topic modelling - automated topic extraction using the latent Dirichlet Distribution (LDA) model (Blei, Ng & Jordan, 2003); second, sentiment analysis (Blei, Ng & Jordan, 2003); and third, topic analysis (Blei, Ng & Jordan, 2003). 2003); second, sentiment analysis - using the VADER model with a sentiment classifier of the BERT pre-

trained model (Hutto & Gilbert, 2014); third, discourse frame identification (frame identification) –

defined with reference to Entman's (1993) operationalisation of frame theory, and reliability was ensured by an inter-coder agreement test (Cohen's kappa ≥ 0.75) (Semetko & Valkenburg, 2000); and fourthly, network analysis - identifying the clustering structure of the discourse community through a modularity algorithm (Newman & Girvan, 2004).

3.3.2 Critical discourse analysis

Based on the macro-statistical results of the computational analysis, this study further selected typical cases for in-depth critical discourse analysis. The cases were selected according to the 'principle of maximum variation' (Patton, 2015). The critical discourse analysis framework adopts Fairclough's (1995, 2003) three-dimensional analysis model, which systematically analyses news discourse in the algorithmic environment from three levels: text, discursive practice and social practice.

3.3.3 Comparative Platform Analysis

This study introduces a comparative dimension to systematically compare the similarities and differences between Twitter/X, Facebook and YouTube in terms of algorithmic gatekeeping mechanisms and the degree of discourse fragmentation. The comparative analysis is centred on the concept of "platform logic" (Bucher & Helmond, 2018), focusing on how the algorithmic design principles, content review standards and user interaction norms of each platform jointly shape the unique discourse ecology of each platform.

3.4 Research Ethics

This study strictly adheres to the ethical norms of academic research, and all public datasets are used in accordance with the terms of use of the original data publisher; social media user data are handled in accordance with the principle of anonymisation; and the research design is in line with the Ethical Guidelines for Research in Digital Media issued by the International Communication Association (ICA) (franzke et al., 2020).

Findings of the study (Results)

4.1 Algorithmic Gatekeeping Content Selection Bias: Emotionalisation and Conflictualisation

4.1.1 Systematic advantage of emotionally-charged content

The results of computational text analysis reveal that among the three types of issues on the three platforms, algorithmic high push content (retweets/likes/shares in the top 25% quartile) is significantly higher than low push content in terms of sentiment intensity indicators. In the case of Twitter political election issues, for example, the mean VADER negative sentiment score was significantly higher for high-tweeted content ($M = -0.42$, $SD = 0.18$) than for low-tweeted content ($M = -0.21$, $SD = 0.15$), and the difference was statistically significant ($t(4,832) = 18.76$, $p < .001$, Cohen's $d = 0.54$). This pattern was similarly replicated for the Facebook climate change topic ($d = 0.61$) and the YouTube new crown epidemic topic ($d = 0.47$) (see Table 1). The above findings are highly consistent with the classic studies by Brady et al. (Brady et al., 2017) and Vosoughi et al. (Vosoughi, Roy & Aral, 2018).

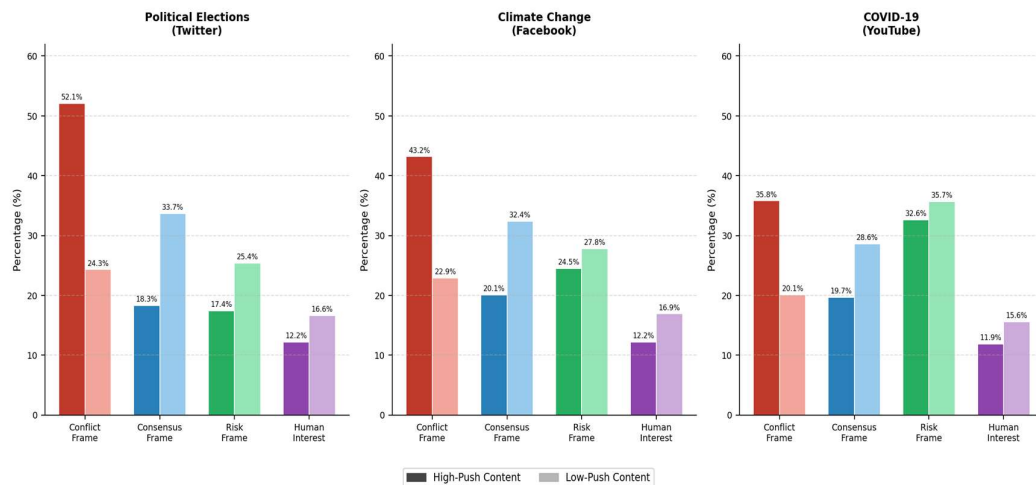
Platform / Issue	Content Type	Mean VADER	SD	Cohen's d	t-statistic (p)
------------------	--------------	------------	----	-----------	-----------------

Algorithmic Gatekeeping And Consensus Fragmentation: How Platform Curation Shapes Public Discourse On Social Media News

		Score				
Twitter – Political Elections	High-Push	-0.42	8	0.1	0.54	t(4832)=18.76, p<.001
	Low-Push	-0.21	5	0.1	—	
Facebook – Climate Change	High-Push	-0.48	1	0.2	0.61	t(3811)=21.43, p<.001
	Low-Push	-0.23	7	0.1	—	
YouTube – COVID-19	High-Push	-0.39	9	0.1	0.47	t(5204)=16.89, p<.001
	Low-Push	-0.20	6	0.1	—	

4.1.2 Dominant strength of conflict frames

The results of the discourse frame analysis show that conflict frames have the highest share of algorithmic high push content, reaching 43.7% of the total post volume on average; in contrast, the combined share of consensus frames and risk frames is higher in algorithmic low push content (about 51.2%) (see Figure The algorithmic dominance of conflict frames is most pronounced on Twitter (52.1% of high tweets), and the symbiotic relationship between ‘risk frames’ and ‘conflict frames’ is most evident in the YouTube comments section (see Table 2).



Platform	Content Type	Conflict Frame (%)	Consensus Frame (%)	Risk Frame (%)	Human Interest (%)
----------	--------------	--------------------	---------------------	----------------	--------------------

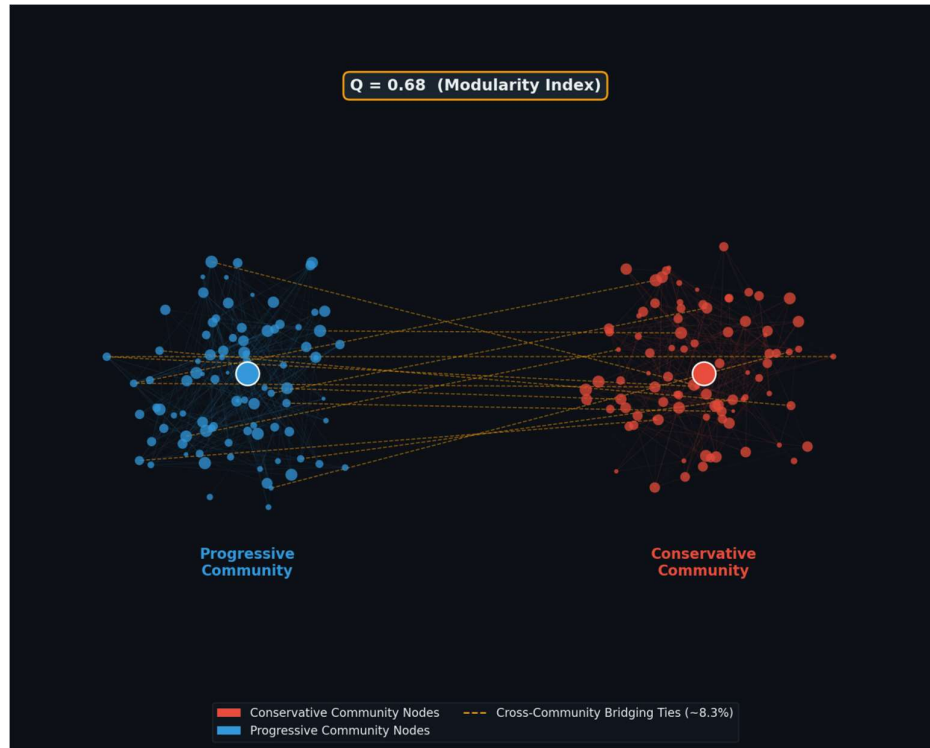
Twitter/ X	High- Push	52.1	18.3	17.4	12.2
	Low- Push	24.3	33.7	25.4	16.6
Faceboo k	High- Push	43.2	20.1	24.5	12.2
	Low- Push	22.9	32.4	27.8	16.9
YouTub e	High- Push	35.8	19.7	32.6	11.9
	Low- Push	20.1	28.6	35.7	15.6

4.2 Quantitative evidence of filtering bubbles: modular structure of discourse networks

4.2.1 High clustering of discourse communities

The results of the network analysis show that the discourse communication networks of all three platforms exhibit a significant modular clustering structure, with modularity indices (Q-values) of $Q = 0.68$ for political election issues on Twitter, $Q = 0.71$ for new crown epidemic issues on Facebook, and $Q = 0.63$ for climate change issues on YouTube, which are much higher than those of $Q = 0.63$ for climate change issues on Twitter as determined by Newman & Girvan (Newman & Girvan, 2004) identified a $Q > 0.3$ significance threshold (see Figure 2). The proportion of bridging ties across communities ranged from only 8.3% (Twitter) to 11.7% (Facebook) of the total number of connections, which is significantly lower than the expected benchmark for a random network (~25%).

Algorithmic Gatekeeping And Consensus
Fragmentation: How Platform Curation Shapes
Public Discourse On Social Media News



4.2.2 Relative Contributions of Algorithmic Pushing vs. Active User Choice

Drawing on the analytical framework of Bakshy et al. (Bakshy, Messing & Adamic, 2015), this study compares the relative contribution of algorithmic effects with user-initiated selection (selective exposure). The results of the analysis show that the algorithmic push mechanism reduces the probability of user exposure to heterogeneous information by about 28.4%, while the contribution of user-initiated choice is about 19.7% (see Table 3). This finding suggests a mutually reinforcing cyclical relationship between algorithmic effects and user-initiated choice (Möller et al., 2018).

Effect Source	Estimated Contribution (%)	95% CI	Operationalisation
Algorithmic Gatekeeping Effect	28.4	[24.7, 32.1]	Reduction in heterogeneous info exposure vs. no-algorithm baseline
User Selective Exposure Effect	19.7	[16.3, 23.1]	Active ideological content avoidance rate
Combined / Interaction Effect	~48.1	[43.2, 53.0]	Total observed homogeneity premium
Note: Estimates based on stratified comparison with			

randomised baseline; methodology adapted from Bakshy et al. (2015).			
---	--	--	--

4.3 Discursive representations of consensus fragmentation: erosion of the cross-group consensus base

4.3.1 Disintegration of the 'common fact base'

The integration of critical discourse analysis and computational analysis reveals a central finding: the 'factual definitions' of the same public issue across different discourse communities are systematically divergent. In the case of the Xinguang epidemic, for example, the 'vaccine risk' frame appears 6.3 times more frequently in conservative-dominated discourse communities than in progressive ones, while the 'system failure' frame appears 4.8 times more frequently in progressive discourse communities than in conservative ones. This "factual divergence" signals the "epistemological crisis" that Lewandowsky et al. (2017) warn of.

4.3.2 Synergistic effects of emotional and cognitive polarisation

The findings also reveal that the fragmentation of discourse facilitated by algorithmic gatekeeping exhibits a synergistic effect of emotional polarisation and cognitive polarisation. Specifically, the frequency of 'de-authorisation' discourse strategy was significantly higher in the

algorithmic high push content than in the low push content (OR = 2.14, 95% CI [1.87, 2.45], $p < .001$), suggesting that the algorithmic mechanism objectively strengthened the communication advantage of anti-agency discourse (see Table 4).

Predictor Variable	OR	95% CI	p-value	Interpretation
Algorithmic High-Push (vs. Low)	2.14	[1.87, 2.45]	<.001	High-push content 2.14× more likely to contain de-authorisation discourse
Platform: Twitter/X (ref: YouTube)	1.73	[1.52, 1.97]	<.001	Twitter context amplifies anti-institutional framing
Platform: Facebook (ref: YouTube)	1.38	[1.21, 1.58]	<.001	Moderate amplification relative to YouTube baseline
Moral-	1.91	[1.74,	<.001	Strong co-

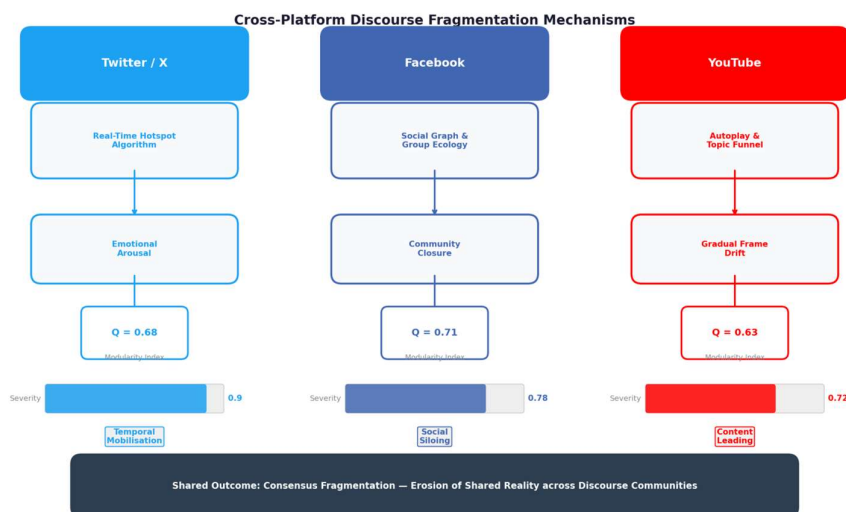
Algorithmic Gatekeeping And Consensus Fragmentation: How Platform Curation Shapes Public Discourse On Social Media News

emotional language intensity		2.10]		occurrence with de-authorisation strategy
Issue type: Political elections (ref: Climate)	1.62	[1.44, 1.82]	<.001	Political context heightens de-authorisation frequency
N = 1,200,000; Nagelkerke R ² = 0.31; Hosmer-Lemeshow $\chi^2(8)$ = 6.84, p = .55				

4.3.3 Cross-platform comparison results

Comparative analyses show that there are significant differences in the degree and mechanism of discourse fragmentation among the three platforms. Overall, Twitter shows the highest degree of emotional arousal and the highest degree of network modularity; Facebook's fragmentation is more reflected in community insularity; and YouTube's fragmentation shows a unique 'topic funnel' phenomenon (see Figure 3). Cinelli et al. (2021) found that YouTube's recommendation algorithm has the strongest unidirectional driving force in guiding users to drift towards extreme content, which is

corroborated by the findings of this study, through a systematic comparison of the echo chamber effect across multiple platforms.



Discussion

5.1 Theoretical Interpretation of the Core Findings: Reconstructing the Framework of 'Algorithmic Public Domain'.

The empirical findings of this study support and deepen the core thesis that algorithmic gatekeeping is not a technology-neutral information filtering tool, but rather a discursive power

mechanism that is deeply political. The findings reveal that the systematic algorithmic dominance of emotionally arousing content (Cohen's d-values ranging from 0.47 to 0.61), the dominance of conflict frames in algorithmically highly pushed content (43.7% on average), and the highly modular structure of discursive communication networks (Q-values of 0.63-0.71) together constitute an "algorithmic public sphere". The structural characteristics that distinguish the 'algorithmic public sphere' from the ideal public sphere.

Habermas (1989) presupposes a space of interactional rationality dominated by the 'uncoercive force of arguments'. However, the findings of this study reveal a diametrically opposed algorithmic logic: the communicative superiority of content depends on its ability to elicit an emotional response rather than on the quality of its arguments (Habermas, 2022). This study therefore proposes the notion of an "algorithmic public sphere", whose core features can be summarised in three points: the algorithmic tyranny of visibility, the structural dominance of emotion, and the institutional erosion of the foundations of consensus.

5.2 Dialogue with established research: support, additions and revisions

5.2.1 Conditional support for filter bubble theory

The findings of this study provide empirical support from multi-platform computational analyses for Pariser's (Pariser, 2011) filter bubble theory, but at the same time make important conditional corrections. This study reveals a mechanistic dimension that has received less attention in the established literature: algorithms systematically reward emotionally arousing content, creating a spiral reinforcement cycle of 'algorithm trains user → user preference reinforces algorithmic logic', which implies that users' so-called 'preferences' are themselves already largely shaped by algorithms over time (Bucher, 2018). Bucher, 2018).

5.2.2 Theoretical additions to political polarisation research

The findings of this study are highly compatible with Iyengar et al.'s (Iyengar et al., 2019) assertion that affective polarisation is increasingly dominating contemporary political discourse, but furthermore reveal a theoretical perspective that goes beyond the framework of political polarisation: algorithmic gatekeeping leads to discursive polarisation not only in terms of ideological polarisation of political views, but also in terms of systematic divergence of 'factual perceptions' of the same public event across different discourse communities. The systematic divergence of different discourse communities' "perceptions" of the same public event is manifested in the fragmentation of epistemological foundations - what Lewandowsky et al. (2017) call the "post-truth society" has gained a new structural support driven by algorithmic mechanisms. support.

5.2.3 Updates to classical gatekeeping theory

This study builds on the multi-level gatekeeping model of Shoemaker & Vos (2009) by proposing that algorithmic gatekeeping introduces a new level of gatekeeping - the platform technical design level. This level is a systemic bias embedded in the platform infrastructure that exerts a continuous and implicit disciplinary force on all content producers through traffic allocation mechanisms (Napoli, 2019). This theoretical addition has important theoretical value for updating gatekeeping theory to fit the platform communication ecology.

5.3 Theoretical implications of platform differences: towards a differentiated platform communication theory

The results of cross-platform comparisons indicate that Twitter/X, Facebook and YouTube differ

significantly in the extent and mechanisms of discourse fragmentation, a finding that has important implications for constructing platform communication theories with greater analytical precision. The findings of this study suggest the need to differentially theorise the algorithmic gatekeeping mechanisms of different platforms in terms of temporality dimension, media modality dimension and business model dimension (Van Dijck et al., 2018), which provides a preliminary analytical framework for the construction of a “differentiated platform discourse theory” (DPT). an initial analytical framework for the construction of differentiated platform discourse theory.

5.4 Policy and practice implications

At the platform governance level, the Digital Services Act (DSA) of the European Union (EU) has been in full force since February 2024, mandating large platforms to conduct transparency disclosure and systematic risk assessment of algorithmic recommendation systems (European Commission, 2022). The findings of this study provide concrete empirical evidence of the “systemic risk” targeted by the DSA, and suggest that the regulatory framework should not adopt a “one-size-fits-all” strategy, but rather differentiate standards for the algorithmic logic of different platforms.

At the level of professional journalism practice, the findings of this study support a practical orientation towards algorithmically-aware journalism (Singer, 2014). At the level of digital media literacy education, the findings highlight the urgency of fostering “algorithmic literacy”, and that incorporating the fundamentals of algorithmic mechanisms into media literacy curricula is an important upstream intervention strategy to address the fragmentation of consensus (Roozenbeek et al., 2020).

5.5 Limitations of the study

There are three main limitations of this study. First, data accessibility limitation: the core parameters of the mainstream platform algorithms are so far highly confidential, and the researcher cannot directly observe the algorithmic decision-making process (Pasquale, 2015). Second, cross-cultural generalisability limitation: the data in this study is dominated by Anglosphere platform content, and algorithmic gate-keeping mechanisms in non-Western contexts may look substantially different depending on differences in platform policies and cultural contexts (Lim, 2023). Third, the limitation of causal inference: this study still faces challenges at the level of strict causal inference, and future studies can establish a more rigorous causal chain of evidence with the help of randomised controlled experiments (Bail et al., 2018).

Conclusion

This study systematically examined how algorithmic gatekeeping mechanisms shape public discourse and lead to consensus fragmentation in social media news dissemination through a triple-method integration of systematic literature synthesis, computational text analysis and critical discourse analysis. The findings make theoretical contributions to established research at three levels.

At the empirical level, this study provides systematic quantitative evidence across platforms and topics, showing that algorithmic gatekeeping mechanisms have a significant push advantage over emotionally charged content and conflicting discourse frames ($d = 0.47-0.61$); the highly modular structure of discourse dissemination networks ($Q = 0.63-0.71$) suggests that cross-community The highly modular structure of the discourse network ($Q = 0.63-0.71$) suggests that cross-group

communication is extremely limited; the systematic divergence of ‘factual definitions’ of the same topic between different discourse communities (with frame frequencies varying on the order of 4-6 times) reveals that fragmentation of the consensus extends beyond affective polarisation to epistemological dimensions.

At the theoretical level, this study proposes the analytical framework of ‘algorithmic public sphere’, locates the discursive politics of algorithmic gatekeeping in the critical inheritance and renewal of Habermas’s theory of the public sphere, adds the key analytical level of ‘platform technology design’ to the traditional theory of gatekeeping, and preliminarily constructs the analytical dimensions of the ‘discursive theory of the differentiated platform’.

At the policy level, the findings support the following policy directions: strengthening the substantive transparency obligations of platform algorithms; promoting discursive diversity as a core normative indicator of platform governance; and incorporating algorithmic literacy education into the systematic framework of digital citizenship cultivation.

Looking ahead, the following topics deserve priority attention: at the level of causal inference, there is an urgent need to establish stricter causal evidence through platform-academic co-operation and randomised experimental design; at the comparative perspective, there is an urgent need for systematic cross-cultural comparative research on algorithmic gate-keeping mechanisms and discourse ecology on non-Western platforms (especially Weibo, WeChat and Shake); at the level of technological evolution, there is an urgent need to study the impact of generative AI on news discourse; and there is an urgent need for the integration of algorithmic literacy education into a systematic digital citizenship cultivation framework. In the dimension of technological evolution, the overlapping effects of generative AI on news discourse production and algorithmic gatekeeping mechanisms will become the next frontier (Napoli, 2019); in the dimension of normative theory, how to redefine the normative

standards of the public sphere in the algorithmic era and translate them into actionable principles of platform design is an important frontier for interdisciplinary co-operation between communication studies and political philosophy.

The systemic impact of algorithmic gatekeeping on public discourse is ultimately a question of power politics about who has the right to decide ‘what is worth being seen’. In an era where the logic of the platform economy commodifies discourse visibility, maintaining a public information ecology that supports democratic consultation and promotes social consensus requires the concerted efforts of academic research, policy regulation, journalistic practice, and civic education. This study hopes to contribute to this academic endeavour..

References

- Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds.). (1998). *Critical realism: Essential readings*. Routledge.
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216-9221. <https://doi.org/10.1073/pnas.1804840115>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132. <https://doi.org/10.1126/science.aaa1160>
- Bhaskar, R. (1998). *The possibility of naturalism: A philosophical critique of the contemporary human sciences* (3rd ed.). Routledge.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3,

Algorithmic Gatekeeping And Consensus
Fragmentation: How Platform Curation Shapes
Public Discourse On Social Media News

993-1022.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313-7318. <https://doi.org/10.1073/pnas.1618923114>

Bucher, T. (2018). *If...then: Algorithmic power and politics*. Oxford University Press. <https://doi.org/10.1093/oso/9780190493028.001.0001>

Bucher, T., & Helmond, A. (2018). The affordances of social media platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE handbook of social media* (pp. 233-253). SAGE.

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociochi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. <https://doi.org/10.1073/pnas.2023301118>

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>

European Commission. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act). *Official Journal of the European Union*.

Fairclough, N. (1995). *Media discourse*. Edward Arnold.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Routledge.

Fisher, M. (2022). *The chaos machine: The inside story of how social media rewired our minds and our world*. Little, Brown and Company.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298-320. <https://doi.org/10.1093/poq/nfw006>

Flick, U. (2018). *An introduction to qualitative research* (6th ed.). SAGE.

franzke, a. s., Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*. Association of Internet Researchers. <https://aoir.org/reports/ethics3.pdf>

Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, (25/26), 56-80. <https://doi.org/10.2307/466240>

Freelon, D. (2015). Discourse architecture, ideology, and democratic norms in online political discussion. *New Media & Society*, 17(5), 772-791. <https://doi.org/10.1177/1461444813513259>

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167-194). MIT Press. <https://doi.org/10.7551/mitpress/9780262525374.003.0009>

Gonzalez-Bailon, S., Lazer, D., Barbera, P., Zhang, M., Allcott, H., Brown, T., ... Tucker, J. A. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, 381(6656), 392-398. <https://doi.org/10.1126/science.ade7138>

Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society* (T. Burger & F. Lawrence, Trans.). MIT Press. (Original work published 1962)

Habermas, J. (2022). *A new structural transformation of the public sphere and deliberative politics* (C. Cronin, Trans.). Polity Press.

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. AAAI Press.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129-146. <https://doi.org/10.1146/annurev-polsci-051117-073034>

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE.

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353-369. <https://doi.org/10.1016/j.jarmac.2017.07.008>

- Lewin, K. (1943). Forces behind food habits and methods of change. *Bulletin of the National Research Council*, 108, 35-65.
- Lim, M. (2023). From activist media to algorithmic politics: The Internet, social media & civil society in Southeast Asia. In E. Hansson & M. Weiss (Eds.), *Routledge handbook of civil and uncivil society in Southeast Asia* (pp. 25-44). Routledge.
- McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176-187. <https://doi.org/10.1086/267990>
- Meraz, S. (2011). Using time series analysis to measure intermedia agenda-setting influence in traditional media and political blog networks. *Journalism & Mass Communication Quarterly*, 88(1), 176-194. <https://doi.org/10.1177/107769901108800110>
- Moller, J., Trilling, D., Helberger, N., & Van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, 21(7), 959-977. <https://doi.org/10.1080/1369118X.2018.1444076>
- Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication Theory*, 24(3), 340-360. <https://doi.org/10.1111/comt.12039>
- Napoli, P. M. (2019). *Social media and the public interest: Media regulation in the disinformation age*. Columbia University Press.
- Nechushtai, E. (2018). Could digital platforms capture the media through infrastructure? *Journalism*, 19(8), 1043-1058. <https://doi.org/10.1177/1464884917725163>
- Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, 90, 298-307. <https://doi.org/10.1016/j.chb.2018.07.043>
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113. <https://doi.org/10.1103/PhysRevE.69.026113>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Papacharissi, Z. (2002). The virtual sphere: The internet as a public sphere. *New Media & Society*, 4(1), 9-27. <https://doi.org/10.1177/14614440222226244>
- Papacharissi, Z. (2015). *Affective publics: Sentiment, technology, and politics*. Oxford University Press.
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE.
- Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16, 101-127. <https://doi.org/10.1146/annurev-polsci-100711-135242>
- Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 across 26 countries. *Royal Society Open Science*, 7(10), 201199. <https://doi.org/10.1098/rsos.201199>
- Semetko, H. A., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, 50(2), 93-109. <https://doi.org/10.1111/j.1460-2466.2000.tb02843.x>
- Shearer, E., & Mitchell, A. (2021). *News use across social media platforms in 2020*. Pew Research Center. <https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/>
- Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory*. Routledge.
- Singer, J. B. (2014). User-generated visibility: Secondary gatekeeping in a shared media space. *New Media & Society*, 16(1), 55-73. <https://doi.org/10.1177/1461444813477833>
- Snelson, C. L. (2016). Qualitative and mixed methods social media research: A review of the literature. *International Journal of Qualitative Methods*, 15(1), 1-15. <https://doi.org/10.1177/1609406915624574>

Algorithmic Gatekeeping And Consensus
Fragmentation: How Platform Curation Shapes
Public Discourse On Social Media News

- Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
- Taneja, H., Webster, J. G., Malthouse, E. C., & Ksiazek, T. B. (2018). Media consumption across platforms: Identifying user-defined repertoires. *New Media & Society*, 14(6), 951-968. <https://doi.org/10.1177/1461444811436146>
- Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press. <https://doi.org/10.1093/oso/9780190889760.001.0001>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. <https://doi.org/10.1126/science.aap9559>
- White, D. M. (1950). The "gate keeper": A case study in the selection of news. *Journalism Quarterly*, 27(4), 383-390. <https://doi.org/10.1177/107769905002700403>
- Zuiderveen Borgesius, F. J., Trilling, D., Moller, J., Bodo, B., De Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, 5(1). <https://doi.org/10.14763/2016.1.401>.