Physica A 665 (2025) 130487

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Polarization in social media☆

Congli Zhang [a,b,c], Xiaolei Wang [a,b], Yong Min [d,e], Shanqing Yu [a,b], Ye Wu [d,e], Qi Xuan [a,b], Chenbo Fu [a,b,*]

[a] The Institute of Cyberspace Security, Zhejiang University of Technology, Hangzhou, 310000, China
[b] Binjiang Institute of Artificial Intelligence, Zhejiang University of Technology, Hangzhou, 310056, China
[c] The School of Engineering, The University of Glasgow, Glasgow, G12 8QQ, United Kingdom
[d] The Computational Communication Research Center, Beijing Normal University, Zhuhai, 519000, China
[e] The School of Journalism and Communication, Beijing Normal University, Beijing, 100000, China

ARTICLE INFO

Keywords:
Polarization quantization
Echo chamber
Social networks
Networks analysis

ABSTRACT

The development of social media has changed the way in which information is consumed by the public. However, it also promotes polarization, especially among the more controversial topics. Furthermore, recommendation systems commonly used in social media have been shown to emerge an echo chamber effect, amplifying user groups' polarization. Most of current studies focus on analyzing the generation of polarization phenomena on social media but rarely investigate how to quantify polarization. In this study, we construct user opinion networks and utilize random walk to quantify the polarization of related topics. The experiments on three real datasets, i.e., Bilibili, YouTube and Reddit, demonstrate that there is polarization in Bilibili and YouTube, especially in Bilibili. Our work complements quantitative measurements of polarization in social media platforms.

1. Introduction

The advent of social media has unprecedentedly changed the mechanism by which we consume information and form opinions [1,2]. Furthermore, users play multiple roles in modern social networks, such as consumers, publishers, and spreaders. This new interactive consumption pattern enables the rapid spread of news and affects the formation of public opinion [3]. Moreover, the massive amount of information in social media has caused information overload for users [4], prompting the platform to develop and utilize recommendation algorithms to help users filter out uninteresting information. These algorithms infer the user's interests through historical data and similarity relationships to render information related to the user's interests [5,6]. Although the recommendation algorithms can enhance the occurrence of information that presents user interests, the filtered content isolates users from the information selection process and limits the user's exposure to non-preferred information [7,8]. Thus, recommendation algorithms may aggregate users with similar interests, forming a virtual space, i.e., echo chambers [9]. In this virtual space, users' opinions, political leanings, or beliefs on a certain topic are strengthened through repeated interactions with other peers, e.g., comments, likes, and forwarding.

☆ This work was supported by the Zhejiang Provincial Natural Science Foundation of China under Grant (ZCLY24F0302), the Zhejiang Fundamental Public Welfare Research Project under Grant (LGF21G010003), the National Natural Science Foundation of China under Grant (62103374 and U21B2001), the Key R&D Program of Zhejiang under Grant (2022C01018), the Social Science Fund of Beijing under Grant (21DTR040), the Key Research and Development Program of Zhejiang Province under Grant (2024C01025), the Fundamental Research Funds for the Central Universities, China under Grant (1243200012), and 2023 General Project of Ministry of Education Foundation on Humanities and Social Sciences under Grant (23YJA860011).

* Corresponding author.
E-mail address: cbfu@zjut.edu.cn (C. Fu).

https://doi.org/10.1016/j.physa.2025.130487
Received 22 August 2024; Received in revised form 19 January 2025

Available online 8 March 2025
0378-4371/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Users in echo chambers tend to interact with users who support their previous opinions, thus forming a user group with convergent opinions. And the more users join this group, the more likely it is to cause polarization [10]. In social science, polarization refers to dividing a society or political group into opposing subgroups with conflicting political standpoints, goals, and perspectives [11]. It is a common phenomenon in social life, such as ideological, emotional, and news audience polarization [4]. For example, in real-world society, the same race people tend to live together, leading to social ethnic polarization [12]. However, unlike the slow formation of real-world polarization, the emergence of echo chambers accelerates the formation of online polarization [13]. In these polarized environments, opposing user groups are prone to communicate in a highly divisive way, with members of the internal groups having only minimal access to external groups and their viewpoints [14,15], which causes the public to begin to worry about the polarization of ideology [16]. For instance, Yardi and Boyd demonstrated that replies between like-minded Twitter users are more frequent than replies between users with different political views, and discussing highly politicized issues tends to strengthen group identity [17]. Adamic and Glance found that political blogs tended to establish connections with other blogs representing similar political ideologies, while they had fewer connections with blogs holding opposite views [18]. Such a highly divisive communication style will affect the public opinion orientation of news on social media platforms [19].

The recommendation algorithms of different social media platforms have promoted the emergence of the echo chamber effect, leading to the widespread polarization of public opinion on the Internet [20–22]. For example, the highly divided partisan report in the US and Japan presidential elections [23,24], discussions on secularism and Islamism during the Egyptian revolution in 2011 [25,26], and the 15M campaign in Spain in 2011 [27]. Especially when it comes to political events, it has been proven to cause social media to split into different communities [28,29]. Schmidt et al. discovered the phenomenon of political polarization of views on vaccine discussions on Facebook [30], while Aman et al. found the phenomenon of emotional polarization and ideological polarization of users on Twitter about climate change [31,32]. Based on user's consumption behavior on YouTube and Facebook, Bessi et al. discovered polarization in science and conspiracy theory consumption among users based on their consumption behavior on YouTube and Facebook [33].

Most current studies primarily focus on modeling polarization dynamics and developing methods for polarization quantification. For instance, Arruda et al. examined how social network algorithms impact polarization [34], while Valensise et al. investigated the drivers of polarization using model fitting [35]. Similarly, Ojer et al. studied depolarization dynamics through simulation [36]. On the quantification side, Yang et al. utilized Twitter content to measure the degree of semantic polarization through word embedding and clustering techniques [37]., but this method is ineffective for non-textual social network platforms, e.g., YouTube. Hohmann et al. used the euclidean distance to quantify the polarization, but it requires high computational costs for large-scale networks [38]. However, quantifying the polarization is still a challenge. In this work, we aim to develop quantitative metrics that systematically measure polarization in user opinion networks. Inspired by recent research [12] on ethnic segregation in metropolitan areas, in this work, we construct user opinion networks on different social media platforms (i.e., Bilibili, YouTube, and Reddit) and utilize the random walk to quantify polarization and explore the differences between different platforms.

Our main contributions are summarized as follows:

1. Construct user opinion networks, discover and quantify polarization on social media platforms through the random walk.
2. Explore the differences in polarization between different social media platforms.
3. Discuss the impact of information factors in social media platforms on polarization.

The rest of this paper is organized as follows: Section 2 introduces related data and the details of our metrics that quantify polarization, Section 3 demonstrates the process of quantifying polarization in specific social media, Section 4 gives detailed experimental results and Section 5 demonstrates our simulation results. Finally, Section 6 concludes our work.

2. Data and methods

2.1. Data collection

The datasets in this work are collected from Bilibili,[1] YouTube[2] and Reddit.[3] For Bilibili and YouTube, we collect video information regarding climate change, feminism, and abortion topics due to their frequent emergence in public discourse and their ability to evoke diverse user opinions. Data collection spanned from 2016 to 2023, ensuring a comprehensive representation of user engagement across these years. Videos were identified through topic-specific keywords (i.e., *Climate Change, COP, Feminism, Female Right, Abortion*). Each video comprises titles, views, likes, user ID, number of comments, and comments content. As for Reddit, since there are few subreddits for climate change, feminism, and abortion, we obtained the data by downloading comments and submissions posted in the subreddit[4] News and Politics. Each blog comprises blog text, number of comments, and comments content. Table 1 shows the detailed information of the dataset.

1 https://www.bilibili.com/.
2 https://www.YouTube.com/.
3 https://www.reddit.com/.
4 https://search.pushshift.io/reddit/.

Table 1
Dataset details.

Media	Topic	#Contents	#Comments	#Users
Bilibili	Climate change	944	14,569	13,122
	Feminism	943	133,498	105,078
	Abortion	923	85,371	72,333
YouTube	Climate change	509	63,043	55,538
	Feminism	563	82,980	75,764
	Abortion	496	101,154	86,085
Reddit	News	1711	24,680,281	48,155
	Politics	4094	47,894,503	178,430

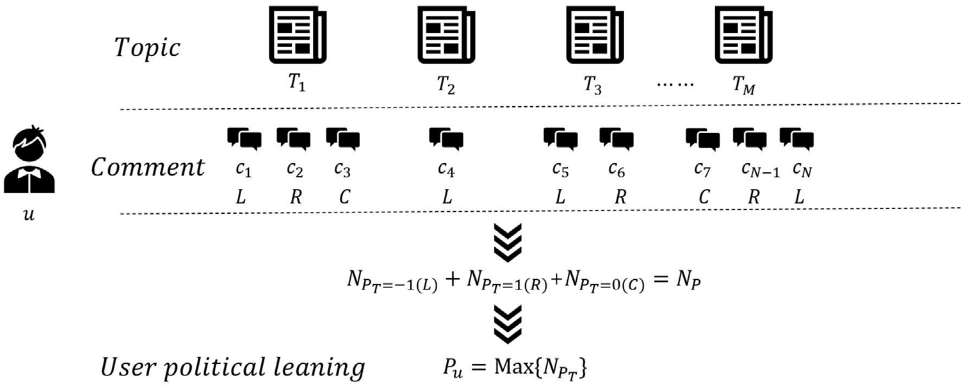

Fig. 1. An illustration of determining user political leaning.

2.2. Methods

2.2.1. Construction of user opinion network

We first determine the user's political leanings toward a topic. Specifically, the individual political leaning of a user u toward a specific topic T can be calculated by the comments the user produces, as the emotion of the comments indirectly reveals the user's political leaning towards the topic [39,40]. Considering there are M numbers of posts related to topic $T = \{T_1, \ldots, T_i, \ldots, T_M\}$ and a user u producing a number N_c of comments, $C_u = \{c_1, \ldots, c_i, \ldots, c_N\}$ towards the topic, and each comment is assigned a political leaning. There are three political leaning states, i.e., $p_T \in [-1, 0, 1]$ associating with left, center, and right-leaning. For example, if a comment supports the post T_i (e.g., the sentiment is positive), then the comment will inherit the political leaning of the post. Furthermore, if the comment opposes the post T_i (e.g., the sentiment is negative), then the comment will inherit the opposite political leaning of T_i. Otherwise, the comment will be assigned as the center if the comment stays neutral about the post. Specifically, we employ TextBlob[5] for English comments and SnowNLP[6] for Chinese comments. Therefore, for an individual post N_p comments toward topic T, there are $N_{p_T=-1}$, $N_{p_T=0}$, and $N_{p_T=1}$ number of leanings associating with three leanings, where $N_{p_T=-1} + N_{p_T=0} + N_{p_T=1} = N$. Thus, an individual's political leaning p_u toward topic T can be regarded as the political leaning that appears most in the content generated by the individual, i.e., $p_u = \max\{N_{p_T}\}$. It should be noted that each user's political leaning is assigned as a single value, categorized by topic. The illustration of calculating user political leanings is shown in Fig. 1.

Next, we construct the user opinion network toward each political topic through user comment relationships to quantify the polarization in users' opinions of Bilibili, YouTube, and Reddit. Specifically, the user opinion network G is an undirected weighted network with N nodes and K edges where nodes represent users, edges represent the co-comment relationship on the same post, and the weight of an edge represents the number of times that co-comments on the same post. It should be noted that we only focus on the largest connected component in our work. The illustration of user opinion network is shown in Fig. 2.

2.2.2. Quantifying polarization through random walk

Inspired by recent research [12] on ethnic segregation in metropolitan areas, we utilize the simplest random walk (which is starting from a node and then transferring to a neighboring node with a transfer probability) to quantify polarization on social media platforms, i.e., the coverage time of the random walk on the corresponding user opinion network. Specifically, our user opinion network contains three groups according to their political leaning states, i.e., left, center, and right-leaning, and the coverage time represents the average expected number of steps required for the walker to visit all political leanings in the network by a random

[5] https://github.com/sloria/TextBlob.
[6] https://github.com/isnowfy/snownlp.

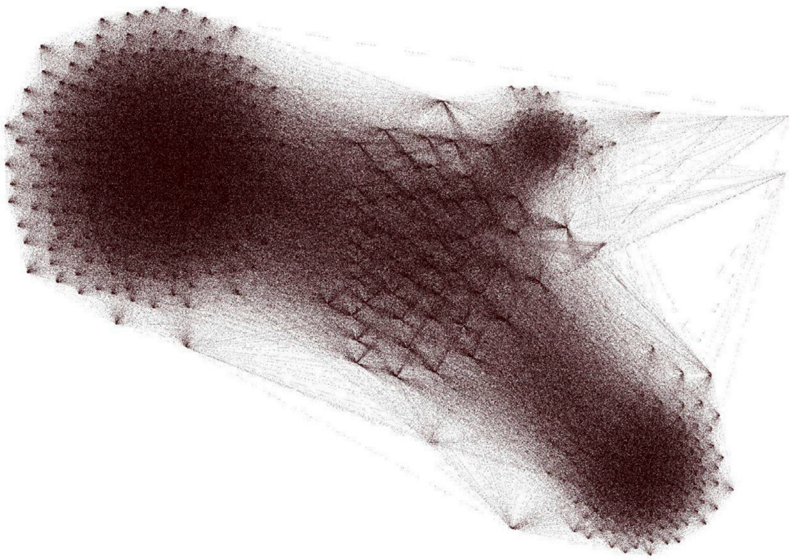

Fig. 2. The illustration of user opinion network (abortion topic on Bilibili).

walk starting from a random initial node i_0 after the R round. In the random walk, the transition probability from node y to node x is calculated as the weight of the edge between the node and its neighbor, i.e., $P(x \mid y) = (w_{yx} - w_{\min})/(w_{\max} - w_{\min})$, if $edge(x, y) \in G$. Otherwise, the transition probability is 0 where w_{yx} represents the weight of the edge between node y and node x, $w_{\min}$ represents the minimum weight of the edge from node y, $w_{\max}$ represents the maximum weight of the edge from node y.

For example, if political leanings are evenly distributed in the user opinion network, the coverage time will not be seriously dependent on the initial node i_0, representing a weak polarization. Conversely, if user nodes tend to form isolated groups with specific political leanings, the coverage time will depend on the initial node, the political leaning of the initial node, the size of other clusters in the network, and the strength of user connections (i.e., the weight of the edge). Generally, the longer the coverage time, the stronger the polarization in the social media platform. In more detail, the average coverage time $S_i(t)$ starting from the node i of a certain topic T after R rounds is shown in the Eq. (1), k represents the number of rounds, where in our work, $R = 1000$.

$$S_i(t) = \frac{1}{R}\sum_{k=1}^{R} S_i \tag{1}$$

For the entire user opinion network, the network coverage time could be the average coverage time of all the nodes, as shown in Eq. (2), where N represents the number of nodes.

$$\mu(t) = \frac{1}{N}\sum_{i=1}^{N} S_i(t) \tag{2}$$

Then, the variance and the diversity of the distribution of coverage time can be calculated as Eqs. (3) and (4), where a_{ij} are the entries of the adjacency matrix of the user opinion network, where K represents the number of edges.

$$\sigma(t) = \frac{\sqrt{\mathrm{Var}\left(S_i(t)\right)}}{\mu(t)} \tag{3}$$

$$\varrho(t) = \frac{1}{K}\sum_{i=1}^{N}\sum_{j<i} a_{ij}\left|S_i(t) - S_j(t)\right| \tag{4}$$

These metrics quantitatively assess polarization, i.e., the user opinion network coverage time $\mu(t)$ characterizes the average coverage time of all the nodes and quantifies the polarization of user groups in the user opinion network; the variance of the distribution of coverage time $\sigma(t)$ quantifies the dependence of the coverage time on the initial node; the diversity of coverage time $\varrho(t)$ quantifies how much the coverage time varies between neighboring nodes in the network.

Table 2
Demonstration of determining video's leaning.

Topic	*Stance*	*Political leaning*
	Support	Left
Climate change	Oppose	Right
	–	Neutral
	Support	Left
Feminism	Oppose	Right
	–	Neutral
	Support	Left
Abortion	Oppose	Right
	–	Neutral

3. Implementation on social media

3.1. YouTube and Bilibili

For social media platforms such as YouTube and Bilibili, where information is spread through videos, we determine user's political leanings according to their comments under videos on related topics. Firstly, the video title determines whether a video post is for or against a certain topic, as shown in Table 2.

Then, the user's attitude toward the video will be used to calculate the user's political leanings toward the video. Specifically, if a user posts a comment under a video and expresses a positive (negative) sentiment, it is construed that the user's contribution aligns with the political leanings endorsed by the video (opposed political leanings for negative sentiment). Otherwise, the user expresses a politically neutral stance. Finally, the political leaning of a user is calculated by counting the political leanings that appear most in user-generated content as mentioned in Section 2.2.1. Furthermore, the user opinion network is constructed by connecting all users who comment on the same video. For example, if user u_i and user u_j comment on the same video, user u_i and user u_j are connected, and the edge weight reflects the frequency with which user u_i and user u_j comment on the same video. In this work, we only focus on the first layer of comments of each video.

3.2. Reddit

For Reddit, where information is conveyed through text, we infer users' political leanings by examining the political orientations of news organizations linked in the blogs users submit. The classification of news organizations' political leanings is based on the MFCB website.[7] For example, if a user publishes a blog with a link to CNN (politically left-leaning media), then the user is deemed to have generated politically left-leaning content. We also count the most frequently occurring political leanings in user-generated content to determine the user's political leanings. The construction of the user opinion network is similar to Bilibili and YouTube, connections are connected between comment users who engage with the blog, e.g., if user u_i and user u_j have commented on the same blog, then user u_i and user u_j are connected, with the edge weight representing the frequency of shared comment of blogs between user u_i and user u_j. Similarly, we only consider the first layer of comments of each blog.

4. Results

To explore platform polarization, we first qualitatively identify polarization in social media. In a polarized environment, users are more likely to be surrounded by users with similar political leanings [41], that is, user i is more likely to be connected to users with similar political leanings. Fig. 3 demonstrates the joint distribution of users' and average neighbors' leanings, where a brighter color indicates greater user density. As shown in Fig. 3(a) and (b), climate change topics on Bilibili demonstrate a strong correlation between the user's political leaning and the average political leaning of their neighbors, followed by YouTube. On YouTube, there is a mix of consensus and polarization according to recent polarization modeling research [34,35]. Contrarily, the Reddit news community does not demonstrate a polarization phenomenon, e.g., there is only a bright area in Fig. 3(c), that is, only containing a single political leaning.

After qualitatively discovering the polarization phenomenon on topics of Bilibili and YouTube, we utilize the random walk proposed in Section 2.2.2 to quantify these phenomena. Fig. 4 demonstrates the normalized coverage time (NCT, i.e., $NCT = (S_i(t) - \mu(t)_{null})/\bar{S}_i(t))$ of the user opinion network in Bilibili, YouTube, and Reddit. $\mu(t)_{null}$ represents the average coverage time of the null model, and the normalized coverage time NCT is used to compare the degree of polarization between the different platforms quantitatively. As shown in Fig. 4, on the Bilibili platform, the median of NCT on abortion, climate change, and female rights topics is larger than 0 and surpasses YouTube, reflecting a comparatively stronger polarization in user discussions. In contrast, the Reddit platform demonstrates a different trend. The median of NCT on these topics is less than 0, suggesting that discussions on political matters within the News and Politics subreddits do not exhibit clear signs of polarization. This absence of polarization

[7] See in https://mediabiasfactcheck.com.

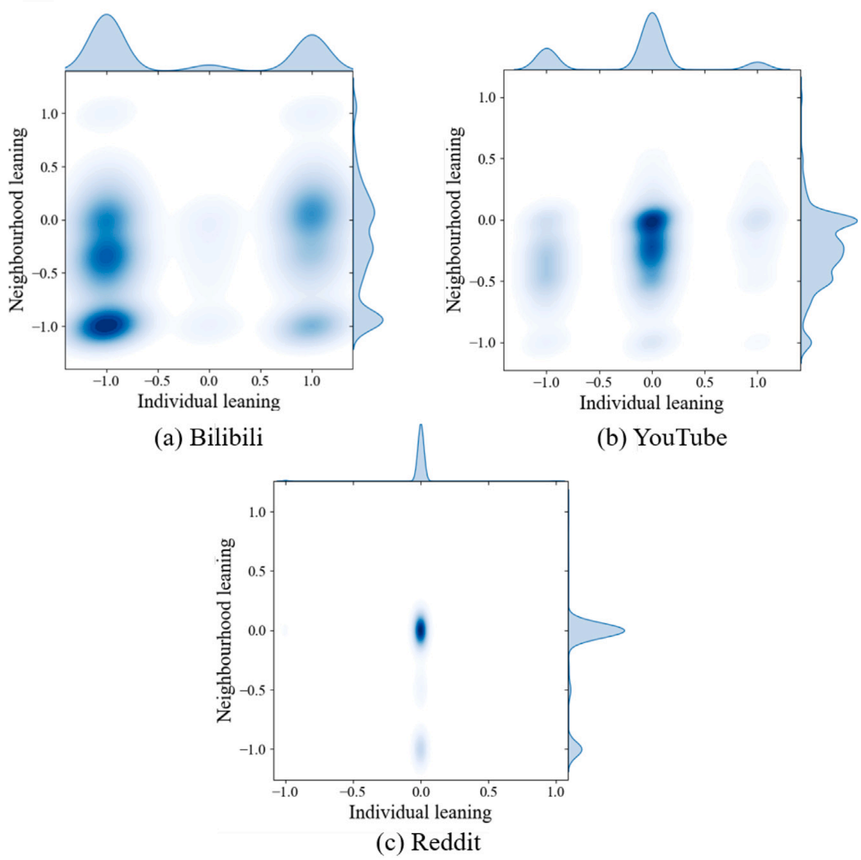

Fig. 3. Joint distribution of users' and average neighbors' leanings for different social media. (a) Bilibili, (b) YouTube, (c) Reddit. where colors represent the density of users and marginal distributions are plotted on the x and y axis, respectively.

Table 3
Kolmogorov–Smirnov test result.

Media	*Topic*	*Statistic*	*p*
Bilibili	*Climate change*	26.3	<0.001
	Feminism	17.2	<0.001
	Abortion	17.9	<0.001
YouTube	*Climate change*	6.65	<0.001
	Feminism	4.69	<0.001
	Abortion	15.0	<0.001
Reddit	*News*	−10.1	<0.001
	Politic	−29.6	<0.001

on Reddit can be attributed to the prevalent left-leaning political orientation among Reddit users [42]. Given this uniform political leaning, there is a lack of confrontation between user groups with differing political inclinations on the platform. Consequently, the existence of polarization is not readily apparent on Reddit. Moreover, we introduce a null model to further verify the polarization phenomenon's existence on Bilibili and YouTube. Specifically, our null model shuffles users' political leanings and edges' weight. Thus, if polarization exists in the user opinion network, the coverage time distributions of the user opinion network and null model are supposed to have significant differences, i.e., the coverage time of the null model is less than the original user opinion network as people with different political leanings are evenly distributed in the network. Hence, we use the Kolmogorov–Smirnov Test to compare the original user opinion network and null model, results show Bilibili and YouTube have statistical differences (Table 3), i.e., $statistics > 0$ and p-value <0.05, implying these platforms contain polarization phenomenon.

Furthermore, we demonstrate three polarization quantification indicators of the user opinion network i.e., $\mu(t)$, $\sigma(t)$ and $\varrho(t)$, on Bilibili and YouTube. The larger the value of $\mu(t)$, the more serious the polarization of user groups in the user opinion network. Similarly, $\sigma(t)$ measures the overall variance of the political leanings coverage time. The larger the value of $\sigma(t)$, the greater the dependence of the political leanings coverage time on the initial node. The larger $\varrho(t)$ represents that the political leanings coverage time of adjacent nodes varies greatly, suggesting that users and neighbors share different political opinions. Furthermore, the larger

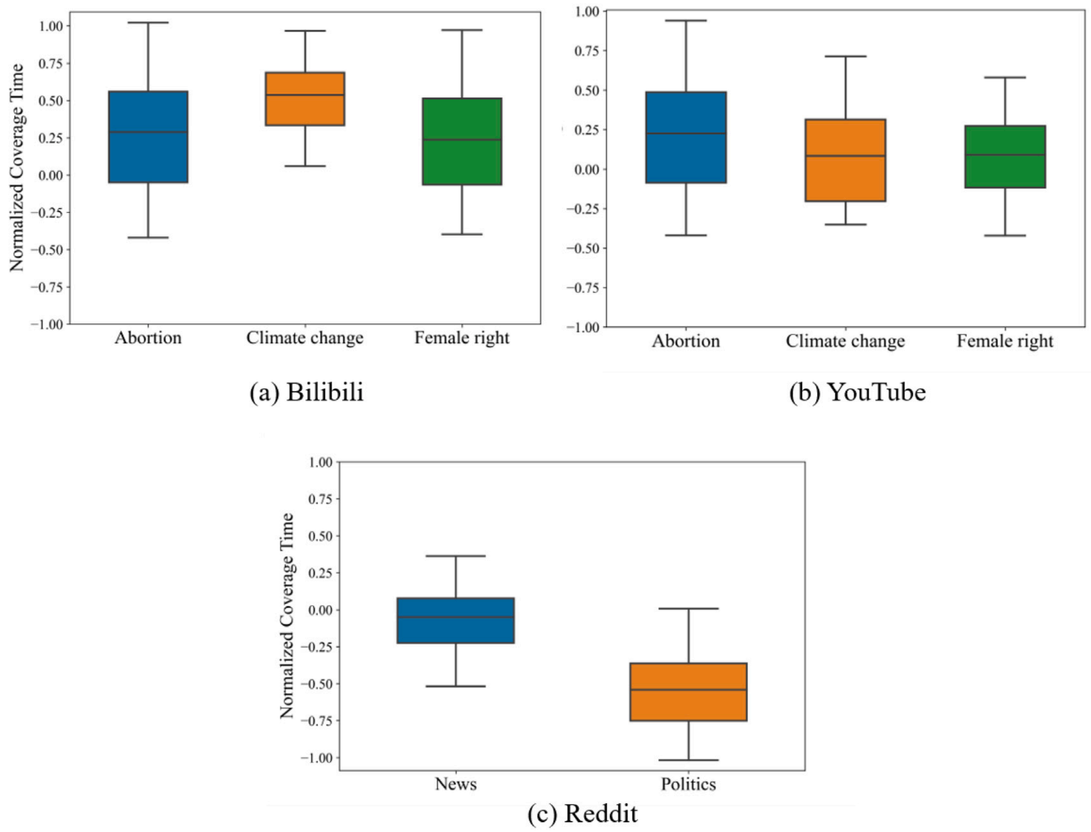

Fig. 4. Normalized coverage time of user opinion network on different topics in (a) Bilibili, (b) YouTube, and (c) Reddit.

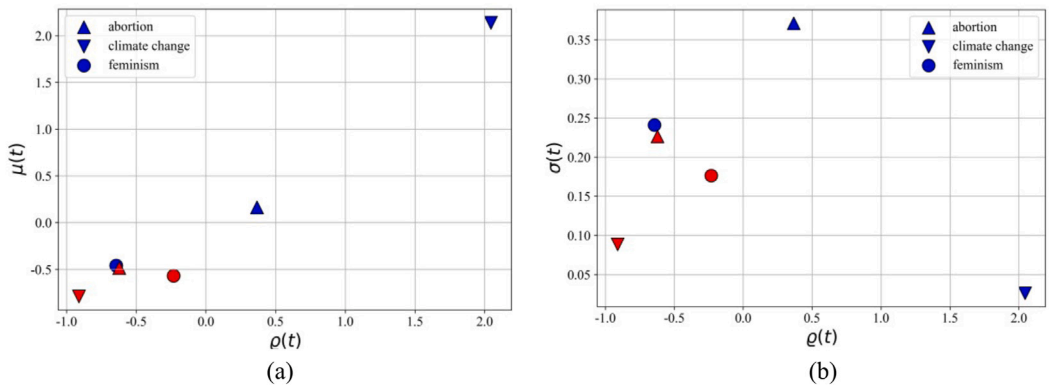

Fig. 5. Polarization indicators of $\mu(t)$, $\sigma(t)$, $\varrho(t)$ in Bilibili (colored in blue) and YouTube (colored in red).

$\varrho(t)$, the stronger the polarization in the network. As shown in Fig. 5, the polarization quantification indicators of these three topics on Bilibili are generally greater than that of YouTube (i.e., in the upper right corner of the figure), which indicates that the degree of polarization of the user discussions about these topics on the Bilibili platform is greater than that of YouTube.

To deeper understand the factors associated with polarization on Bilibili and YouTube, we use the correlation t-test to explore the correlation between topic features (i.e., average number of video views, average number of fans, and average number of comments) and polarization quantification indicators (i.e., $\varrho(t)$, $\sigma(t)$, $\mu(t)$), as shown in Table 4, where * represents $p < 0.05$. The results demonstrate that $\varrho(t)$ and $\mu(t)$ have a significant negative correlation with the number of video comments, suggesting that the topics with more comments, the more difficult it is to generate polarization. This may be because as the topic is widely discussed, users will continue to deepen their thinking about it and become more inclusive, leading to less user polarization.

Table 4
The relationship between average topic features and polarization quantification indicators.

Average topic features	t-test (p-value)		
	$\rho(t)$	$\sigma(t)$	$\mu(t)$
#views	−0.712(0.287)	−0.183(0.816)	−0.720(0.280)
#thumbs	−0.734(0.266)	−0.182(0.818)	−0.738(0.262)
#fans	−0.527(0.473)	−0.352(0.648)	−0.531(0.469)
#comments	−0.985(0.015*)	0.333(0.667)	−0.973(0.027*)

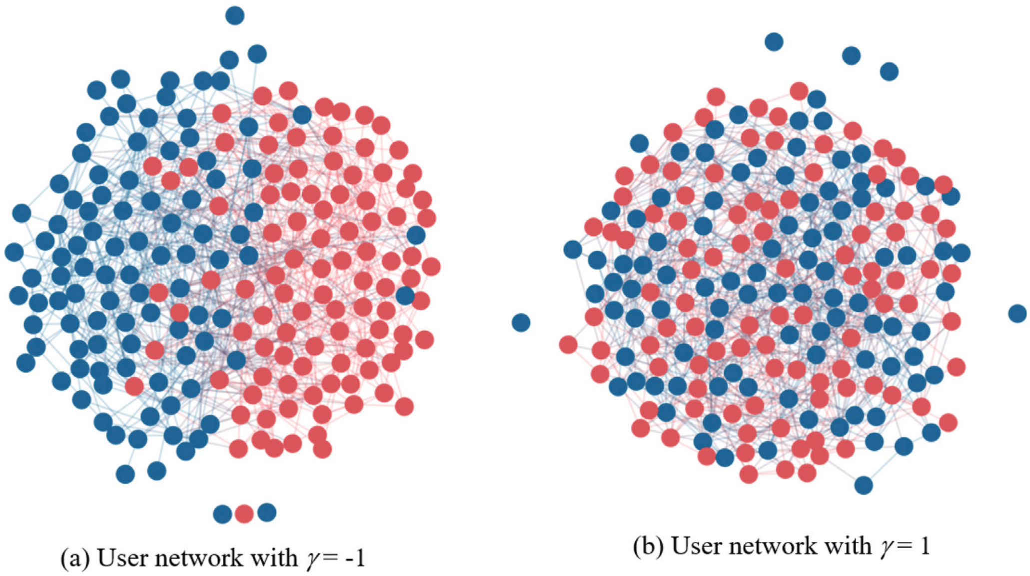

Fig. 6. The relationship between Normalized coverage time and polarized ($\gamma = -1.0$) and non-polarized ($\gamma = 1.0$) social networks.

5. Simulation

To verify the effectiveness of our proposed method, we conduct a virtual user network (contains N nodes and K edges) with political leaning. Our work uses the Tokita model [43] to simulate a polarized social network, where nodes represent users and edges represent social relationships. For simplicity, in our simulation, we only consider left and right political leaning, i.e., each user has a fixed political leaning, either L or R, where in this work, $N = 200$ and $K = 800$.

To generate a polarized social network, we first introduce two information sources, i.e., M_L and M_R, which are associated with L and R leanings. Then, in each evaluation round, social connections change through two steps. In the first step, we filter activated users who are influenced by themselves or surrounding neighborhoods. Specifically, we randomly selected 10% users to sample their corresponding media sources. These users respond based on a news sensitivity θ_i: they become activated if $s > \theta_i$. Otherwise, they remain inactive. The remaining 90% users' status are influenced by the activity of their neighbors, e.g., an individual becomes active if the fraction of their active neighbors exceeds their threshold. Once a user becomes active, they remain in the active state for the rest of the round where the significance ($s \in [0, 1]$) of users is used to measure the difference between two sources reporting on the same story, e.g., Left-leaning or Right-leaning. Since the content of coverage varies from different media sources to the same topic [44]. The significance (s) is simulated by drawing the significance s_L of news reporting from sources M_L and the significance s_R of news reporting from sources M_R from a multivariate normal distribution $N(0, \Sigma)$, where Σ is the covariance matrix with entries $\Sigma_{12} = \Sigma_{21} = \gamma$ and $\Sigma_{11} = \Sigma_{22} = 1$. s_L and s_R are then normalized to the interval $(0, 1)$ using a cumulative distribution function for gaussian distribution. $\gamma \in [-1, 1]$ represents the correlation between the two media sources. A high γ (close to 1.0) implies a similar report, while a low γ (close to -1.0) indicates dissimilarity. In the second step, activated users adjust their social relationships. In this step, we randomly select users with an activated state. If the user is influenced by the activity of their neighbors, she randomly breaks a social tie with an activated neighbor. It should be noted that, in our work, the total number of social relationships is kept to prevent networks from irreversibly fragmenting. Thus, if the social tie is broken, a new social tie is added between two unrelated randomly chosen users.

In our simulation, we set $\gamma = -1.0$ and $\gamma = 1.0$ and obtained two user networks as shown in Fig. 6(a), where the blue nodes represent users with political leaning R and the red nodes represent with political leaning L. To measure the political polarization, we calculated the assortativity of two networks, which is the tendency of nodes to attach to other nodes that are in the same political leanings [45]. This metric takes values in the range $[-1, 1]$, with positive values indicating that individuals tend to be attached to

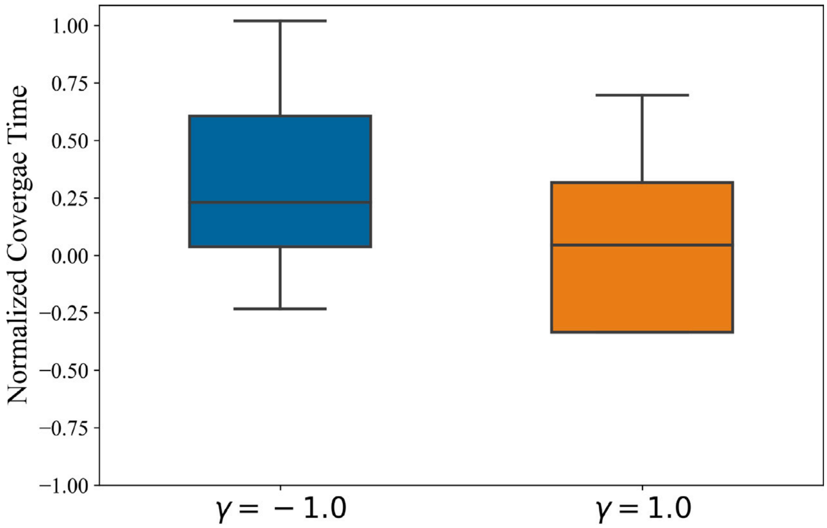

Fig. 7. The relationship between Normalized coverage time and polarized ($\gamma = -1.0$) and non-polarized ($\gamma = 1.0$) social networks.

others of the same political leaning and negative values indicating that individuals tend to be attached to others of opposite political leanings. The political assortativity of the user network when $\gamma = -1.0$ is 0.34 and 0.0 while $\gamma = 1.0$. Then, we use random walk to quantify the polarization of the user network simulated when $\gamma = -1.0$ (polarized social network) and $\gamma = 1.0$ (unpolarized social network). The coverage time of the user network when $\gamma = -1.0$ is larger than 0, demonstrating a clear polarization as shown in Fig. 7, which verifies the feasibility of our quantitative polarization method.

6. Conclusion

To summarize, our present work proposed a new polarization quantification indicator that can efficiently quantify the degree of polarization in social media platforms. Unlike measuring controversy [46], our primary goal is to quantify polarization, which emphasizes the segregation and clustering of users with similar leanings rather than the contentious interactions between groups. In the context of polarization, a highly polarized network is characterized by strong intra-group connections (users with similar leanings) and weak inter-group connections (limited interaction between users with opposing leanings). As a result, in a polarized network, it takes significantly longer for a random walker to move between groups, leading to a higher coverage time. Conversely, in less polarized networks, where cross-group interactions are more frequent, the coverage time is lower. Specifically, by utilizing the coverage time of user opinion networks, we quantify the polarization in three real social networks, i.e., Bilibili, YouTube, and Reddit. The results show that there exist significant distinctions between different social media platforms, e.g., Bilibili and YouTube demonstrate significant user polarization, and in contrast, Reddit does not demonstrate significant user polarization. Furthermore, we observed differences in user polarization between platforms in discussions on polarization topics. For example, the degree of user polarization for discussions about abortion and climate change in Bilibili is greater than that on YouTube, while the degree of user polarization for discussions about feminism is less than that on YouTube. Moreover, the correlation analysis demonstrates that topics with more comments by users are less likely to be polarized, though the small sample size may affect the reliability of the Pearson correlation results.

However, our work does not discuss the evolution of polarization over time. Thus, in the future, we will explore the development of polarization in social media platforms to understand better how different recommendation algorithms influence polarization formation. Hopefully, our findings could be further expanded to uncover social dynamics and information consumption on social media and help social media platform administrators develop policies to reduce polarization on platforms.

CRediT authorship contribution statement

Congli Zhang: Writing – review & editing, Writing – original draft, Software, Investigation, Formal analysis, Data curation. **Xiaolei Wang:** Writing – review & editing, Visualization, Validation. **Yong Min:** Writing – review & editing, Investigation, Conceptualization. **Shanqing Yu:** Writing – review & editing, Validation, Funding acquisition. **Ye Wu:** Writing – review & editing, Conceptualization. **Qi Xuan:** Writing – review & editing, Supervision, Funding acquisition. **Chenbo Fu:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Zhejiang Provincial Natural Science Foundation of China under Grant (ZCLY24F0302), the Zhejiang Fundamental Public Welfare Research Project under Grant (LGF21G010003), the National Natural Science Foundation of China under Grant (62103374 and U21B2001), the Key R&D Program of Zhejiang under Grant (2022C01018), the Social Science Fund of Beijing under Grant (21DTR040), the Key Research and Development Program of Zhejiang Province under Grant (2024C01025), the Fundamental Research Funds for the Central Universities, China under Grant (1243200012), and 2023 General Project of Ministry of Education Foundation on Humanities and Social Sciences under Grant (23YJA860011).)

Data availability

Data will be made available on request.

References

[1] S. Flaxman, S. Goel, J.M. Rao, Filter bubbles, echo chambers, and online news consumption, Public Opin. Q. 80 (S1) (2016) 298–320.
[2] L. Bode, Political news in the news feed: Learning politics from social media, Mass Commun. Soc. 19 (1) (2016) 24–48.
[3] L. Terren, Open University of Catalonia, R. Borge-Bravo, Open University of Catalonia, echo chambers on social media: A systematic review of the literature, Rev. Commun. Res. 9 (2021) 99–118.
[4] A. Ross Arguedas, C. Robertson, R. Fletcher, R. Nielsen, Echo chambers, filter bubbles, and polarisation: A literature review, 2022.
[5] M. Haim, A. Graefe, H.-B. Brosius, Burst of the filter bubble? Effects of personalization on the diversity of google news, Digit. Journal. 6 (3) (2018) 330–343.
[6] C. Fu, X. Yue, B. Shen, S. Yu, Y. Min, Patterns of interest change in stack overflow, Sci. Rep. 12 (1) (2022) 11466.
[7] E. Bakshy, S. Messing, L.A. Adamic, Exposure to ideologically diverse news and opinion on facebook, Science 348 (6239) (2015) 1130–1132.
[8] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C.M. Valensise, E. Brugnoli, A.L. Schmidt, P. Zola, F. Zollo, A. Scala, The covid-19 social media infodemic, Sci. Rep. 10 (1) (2020) 1–10.
[9] K.H. Jamieson, J.N. Cappella, Echo Chamber: Rush Limbaugh and the Conservative Media Establishment, Oxford University Press, 2008.
[10] J.J. Van Bavel, S. Rathje, E. Harris, C. Robertson, A. Sternisko, How social media shapes polarization, Trends Cogn. Sci. 25 (11) (2021) 913–916.
[11] L.G. Smith, E.F. Thomas, A.-M. Bliuc, C. McGarty, Polarization is the psychological foundation of collective engagement, Commun. Psychol. 2 (1) (2024) 41.
[12] S. Sousa, V. Nicosia, Quantifying ethnic segregation in cities through random walks, Nat. Commun. 13 (1) (2022) 5809.
[13] F. Baumann, P. Lorenz-Spreen, I.M. Sokolov, M. Starnini, Modeling echo chambers and polarization dynamics in social networks, Phys. Rev. Lett. 124 (4) (2020) 048301.
[14] C.R. Sunstein, The Law of Group Polarization, John M. Olin Law & Economics Working Paper (91), University of Chicago Law School, 1999.
[15] D.J. Isenberg, Group polarization: A critical review and meta-analysis, J. Pers. Soc. Psychol. 50 (6) (1986) 1141.
[16] E. Kubin, C. von Sikorski, The role of (social) media in political polarization: a systematic review, Ann. Int. Commun. Assoc. 45 (3) (2021) 188–206.
[17] S. Yardi, D. Boyd, Dynamic debates: An analysis of group polarization over time on twitter, Bull. Sci. Technol. Soc. 30 (5) (2010) 316–327.
[18] L.A. Adamic, N. Glance, The political blogosphere and the 2004 us election: divided they blog, in: Proceedings of the 3rd International Workshop on Link Discovery, 2005, pp. 36–43.
[19] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (6380) (2018) 1146–1151.
[20] G.J. Martin, J. McCrain, Local news and national politics, Am. Political Sci. Rev. 113 (2) (2019) 372–384.
[21] M. Levendusky, N. Malhotra, Does media coverage of partisan polarization affect political attitudes? Political Commun. 33 (2) (2016) 283–301.
[22] M.S. Levendusky, Why do partisan media polarize viewers? Am. J. Political Sci. 57 (3) (2013) 611–623.
[23] A. Hanna, C. Wells, P. Maurer, L. Friedland, D. Shah, J. Matthes, Partisan alignments and political polarization online: A computational approach to understanding the french and us presidential elections, in: Proceedings of the 2nd Workshop on Politics, Elections and Data, 2013, pp. 15–22.
[24] S. Usui, M. Yoshida, F. Toriumi, Analysis of information polarization during Japan's 2017 election, in: 2018 IEEE International Conference on Big Data, Big Data, IEEE, Seattle, WA, USA, 2018, pp. 4383–4386.
[25] I. Weber, V.R.K. Garimella, A. Batayneh, Secular vs. islamist polarization in egypt on twitter, in: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013, pp. 290–297.
[26] J. Borge-Holthoefer, W. Magdy, K. Darwish, I. Weber, Content and network dynamics behind egyptian political polarization on twitter, in: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 2015, pp. 700–711.
[27] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, Y. Moreno, The dynamics of protest recruitment through an online network, Sci. Rep. 1 (1) (2011) 1–7.
[28] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, A. Flammini, Political polarization on twitter, in: Proceedings of the International Aaai Conference on Web and Social Media, vol. 5, 2011, pp. 89–96.
[29] M.D. Conover, B. Gonçalves, A. Flammini, F. Menczer, Partisan asymmetries in online political activity, EPJ Data Sci. 1 (1) (2012) 1–19.
[30] A.L. Schmidt, F. Zollo, A. Scala, C. Betsch, W. Quattrociocchi, Polarization of the vaccination debate on Facebook, Vaccine 36 (25) (2018) 3606–3612.
[31] A. Tyagi, J. Uyheng, K.M. Carley, Affective polarization in online climate change discourse on Twitter, in: 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM, IEEE, The Hague, Netherlands, 2020, pp. 443–447.
[32] M. Falkenberg, A. Galeazzi, M. Torricelli, N. Di Marco, F. Larosa, M. Sas, A. Mekacher, W. Pearce, F. Zollo, W. Quattrociocchi, et al., Growing polarization around climate change on social media, Nature Clim. Change (2022) 1–8.
[33] A. Bessi, F. Zollo, M. Del Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, W. Quattrociocchi, Users polarization on Facebook and Youtube, PLoS One 11 (8) (2016) e0159641.
[34] H.F. de Arruda, F.M. Cardoso, G.F. de Arruda, A.R. Hernández, L. da Fontoura Costa, Y. Moreno, Modelling how social network algorithms can influence opinion polarization, Inform. Sci. 588 (2022) 265–278.

[35] C.M. Valensise, M. Cinelli, W. Quattrociocchi, The drivers of online polarization: Fitting models to data, Inform. Sci. 642 (2023) 119152.
[36] J. Ojer, M. Starnini, R. Pastor-Satorras, Modeling explosive opinion depolarization in interdependent topics, Phys. Rev. Lett. 130 (20) (2023) 207401.
[37] M. Yang, X. Wen, Y.-R. Lin, L. Deng, Quantifying content polarization on Twitter, in: 2017 IEEE 3rd International Conference on Collaboration and Internet Computing, CIC, IEEE, San Jose, CA, 2017, pp. 299–308.
[38] M. Hohmann, K. Devriendt, M. Coscia, Quantifying ideological polarization on a network using generalized Euclidean distance, Sci. Adv. 9 (9) (2023) eabq2044.
[39] L. Wang, K. Kim, Analyzing group polarization through text emotion measurement and time series prediction: A comparative study across three online platforms, Measurement: Sensors (2024) 101216.
[40] L. Xie, D. Wang, F. Ma, Analysis of individual characteristics influencing user polarization in covid-19 vaccine hesitancy, Comput. Hum. Behav. 143 (2023) 107649.
[41] W. Cota, S.C. Ferreira, R. Pastor-Satorras, M. Starnini, Quantifying echo chamber effects in information spreading over political communication networks, EPJ Data Sci. 8 (1) (2019) 35.
[42] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, The echo chamber effect on social media, Proc. Natl. Acad. Sci. 118 (9) (2021) e2023301118.
[43] C.K. Tokita, A.M. Guess, C.E. Tarnita, Polarized information ecosystems can reorganize social networks via information cascades, Proc. Natl. Acad. Sci. 118 (50) (2021) e2102147118.
[44] R. Puglisi, J.M. Snyder Jr., Newspaper coverage of political scandals, J. Politics 73 (3) (2011) 931–950.
[45] M.E. Newman, Mixing patterns in networks, Phys. Rev. E 67 (2) (2003) 026126.
[46] K. Garimella, G.D.F. Morales, A. Gionis, M. Mathioudakis, Quantifying controversy on social media, ACM Trans. Soc. Comput. 1 (1) (2018) 1–27.