Proceedings of the 2nd International Conference on Social Psychology and Humanity Studies
DOI: 10.54254/2753-7048/37/20240547

Strategies for Controlling Negative Content on Social Media: A Game Theoretical Perspective

Yan Li[1,a,*]

[1]*School of Economics, Renmin University of China, Beijing, 10010, China*
a. 20203929@stu.hebmu.edu.cn
**corresponding author*

Abstract: With the increase of social network users, using social media platforms has become an indispensable part of every individual's life, but the explosive growth in the amount of information has also increased the difficulty of governing detrimental content on social media platforms. Maintaining a good network environment requires not only proper guidance from the government but also the joint efforts of social media platforms and online users. This paper innovatively divides network users into information producers and information consumers and establishes a tripartite game model with social media platforms. This paper aimed to find out the important factors that are affecting the dissemination of detrimental information by analyzing the changes in equilibrium. The paper suggests that the government can directly influence the operation of social media platforms by adopting different policies, and it is important to grasp the strictness of the policies. Secondly, ordinary social media users are a force to be reckoned with in the process of governing negative information. Finally, what kind of content an information producer publishes depends not only on the revenue but also on the risk he will face. Based on the above findings, this paper offers suggestions on the governance of bad information from the perspectives of the government and the three game players.

Keywords: Social media platforms, Platform governance, Tripartite game

1. Introduction

With the number of global social network users approaching 5 billion and the increasing decentralization, social media platforms have become the main channels for people to communicate, share information, and build a global social network [1]. Social media platforms are easy to use so that every user can become a protagonist on the Internet. But this also makes social media platforms a perfect place where detrimental information can spread quickly. Detrimental information includes false information, vulgar information, hate speech, and other potentially harmful content. They permeate every corner of the Internet, not only polluting the network environment but also bringing a series of problems, such as leakage of personal information, online fraud, cyber violence, social antagonisms, conflicts, etc., which cause great losses to both individuals and society.

Social media network consists of social media platforms and users. The platform is the channel of information dissemination, where users share and understand different views. The users are the producers and consumers of information, as well as the disseminators of information. The average social media user spends close to 2.5 hours per day on social media [1]. During this time users are

DOI: 10.54254/2753-7048/37/20240547

producing and consuming information in different forms and contents, such as video, audio, images, text, etc. And the quantity is also beyond our imagination, which makes it more difficult to control detrimental information.

At present, the more mature text, picture, video, and audio recognition technology are the k-nearest neighbors (KNN) algorithm, Naive Bayes (NB) algorithm, message-digest algorithm 5(MD5), Optical Character Recognition (OCR), and so on. Although these techniques can filter detrimental information in different situations, they have problems such as misjudgment and slow recognition speed [2]. With the application of artificial intelligence technology and deep learning technology in a wide range of fields, the technology for managing detrimental information has been further developed. With the application of artificial intelligence technology and deep learning technology in a wide range of fields, the technology for managing detrimental information has been further developed. For example, Liang Yanliang et al proposed a technical solution for Internet Data Center (IDC) content monitoring system based on artificial intelligence that helps to improve network security and network environment [3]. In addition, the TextCNN-Bert fusion model, text filtering model based on BP neural network, and artificial neural network have also been proven to be effective in recognizing and filtering detrimental information [4-6]. However, due to the cost of time, technological limitations lead to the fact that these natural language processing techniques are also inseparable from the process of manual review [7]. Considering the current climate where online traffic equates to profitability, major social media platforms often exhibit reluctance towards technological advancements that could curb the spread of harmful information.

2. Literature Review

Game theory has obvious advantages in studying the behavioral strategies between multiple subjects of interest. Therefore, although the choice of game subjects is different, many scholars choose to establish evolutionary game models to analyze problems related to the governance of detrimental information.

Lin Ling and Chen Fuji selected four key stakeholders for their study: internet influencers, social media networks, governmental bodies, and platform users. They analyzed that strengthening the government's supervision and improving the information recognition ability of network users are conducive to purifying the network environment. They also emphasized the important position of authoritative media in the dissemination of public opinion [8]. Wang Renda and Li Haifen separated the monitoring department from the government and constructed a tripartite evolutionary game among the government, social media platforms, and the monitoring department. They emphasized the importance of the government's macro-control in governing detrimental information. They also proposed to ensure the autonomy of the monitoring department [9]. Zhang Bin et al and Sun Xiaoyang all chose three subjects such as the government, social media platforms, and users, but the former constructed a three-party game while the latter constructed a two-by-two game [10-11]. Wang Yiming et al demonstrated that the audit incentive mechanism can effectively improve the quality of social media user interaction by constructing a game between social media platforms and users [12]. Luo Mengying et al introduced the concept of an unconfirmed information purifier. They constructed a game model between purifier and social media platforms and emphasized that social media platforms play a major role in the governance of detrimental information [13].

In addition, Zhang Yuying discusses the power contrast between the government, private sector, international non-governmental organizations, and individuals in the governance process of social media platforms. He proposes that different subjects with different interests should jointly participate in the governance of social media platforms and form a reasonable and effective regulatory system [14]. Mo Zuying compared the differences that exist in the static game of incomplete information between information providers and information users when information managers perform different

DOI: 10.54254/2753-7048/37/20240547

functions. He proposed that information managers can improve the quality of information resources by adjusting the incentives given to information providers [15]. Li Gang and Song Qiang established three static game models with complete information between the government and content providers, platform providers, and users. He suggested to effectively control the negative effects of government regulation. He proposed to reduce the exposure rate of detrimental information in various ways and increase the cost of posting detrimental information by Internet users [16].

To summarize, previous studies have shown diversity in the selection of the main body of detrimental information governance, but there was a lack of research methods. Most of the studies have been analyzed by constructing an evolutionary game model and carrying out system simulation. Second, the analyses involving the government have emphasized to varying degrees the importance of the government as a regulator in governing information. No matter the government's macro-control or the punishment system and subsidy system for social media platforms is easy to change the equilibrium state. However, compared to the government, scholars have limited discussion on Internet users. Previous studies did not show the difference between network users as information producers and users as information consumers in governing detrimental information.

Based on the above, this paper will construct a complete information static game among social media platforms, information producers, and information consumers. This paper will summarize the important factors in governing detrimental information in social media by discussing the equilibrium strategies in different situations and making relevant suggestions.

3. Game Modeling and Equilibrium Analysis

3.1. Game Players

The network environment of social media platforms is closely related to platforms, information producers, and information consumers. As a channel for information dissemination, platforms must audit the information on their platforms. If the platform's inadequate management leads to large economic losses for users or triggers a bad impact on society, it will face the risk of being penalized by government regulators. But at the same time social media platforms as an enterprise, it is profit oriented. Therefore, it needs to reduce the cost of auditing and utilize diverse information to expand social influence and attract more platform users to gain traffic. Information producers are the publishers of diverse information, who generally utilize the attention of platform users and their influence on the network to gain financial benefits, so what kind of content they choose is crucial to them. Information consumers are the widest range of ordinary users of social media platforms, and they are the recipients of information. They are the recipients of information. In the face of a wide variety of information, they must improve their ability to recognize information so that they do not fall into the information trap.

3.2. Model Assumptions

(1) The subjects of the game are rational, and all seek to obtain the maximum benefit [17]. This is the common knowledge of the participants [18].

(2) The social media platform's strategy set S_1= {high-intensity management, low-intensity management}. "High-intensity management" means that the platform uses the latest information filtering technology to actively review the information on the platform. At this time, the platform will provide traffic and higher revenue share to high-quality information and its publishers, while deleting, limiting traffic, and blocking detrimental information and its publishers. "Low-intensity management" means that the platform does not actively review information on its platform, but only handles complaints from information consumers.

DOI: 10.54254/2753-7048/37/20240547

The information producer's strategy set S_2 = {self-regulating, non-self-regulating}. "Self-regulating" means that the information producer will self-censor when posting information and will not post videos or articles that contain detrimental information. It also means that information producer actively manages their channels to maintain a healthy network environment. "Non-self-regulating" means that the information producer publishes exaggerated and false information to attract the attention of Internet users.

The information consumer's strategy set S_3= {participatory, non-participatory}. "Participatory" means that information consumer will improve their ability to identify different information and participate in the management of detrimental information by reporting detrimental information, supporting high-quality information, or using other social media platforms. "Non-participatory" means that the information consumer does not identify information on his own and does not participate in the management of detrimental information.

(3) Assuming that every benefit and cost associated with the social media platform, information producer and information consumer are quantifiable and non-negative

When an information consumer browses information, the social media platform gains benefit A_1, and the information producer gains benefit A_2. When an information consumer browses content that does not contain detrimental information (high-quality information), the information consumer receives a revenue of A_3. When the information producer increases the credibility of the social media platform by posting high-quality information, the platform receives an additional benefit B_1. When an information consumer browses high-quality information and increases the influence of information producers by commenting, following, sharing, etc, the information producer receives additional benefit B_2. When the platform adopts the latest information filtering technology, the government provides subsidy B_3.

The platform that chooses the strategy "high-intensity management" needs to introduce new technologies and employ more people to review information, thus incurring an additional cost C_1 compared to chooses the strategy "low-intensity management". Information producer who chooses the strategy "self-regulating" incur an additional cost C_2 compared to those choosing the strategy "non-self-regulating". Because the information producer needs to spend more time controlling the quality and content of information. Information consumer participation in governance requires additional time and effort to identify information, thus incurring an additional cost C_3 compared to non-participation in governance. If an information consumer participates in governance and finds that the platform has no available information or that detrimental information exists, he will choose to stop using the platform. The loss to the platform caused by lost a user is D_1. Detrimental information generates damages to information consumer who is unable to recognize the harmful information as D_2. When the platform's network environment is bad, the platform receives a penalty of D_3 from the government.

Based on the above assumptions, the payoff matrix of the tripartite game is shown in Table 1.

Table 1: Payoff matrix of the tripartite game between social media platform, information producer, and information consumer.

	Information Consumer: participatory		Information Consumer: non-participatory	
	Information Producer			
	Self-regulating	non-self-regulating	self-regulating	Inon-self-regulating

DOI: 10.54254/2753-7048/37/20240547

Table 1: (continued)

SM Platform: high-intensity management	A_1-C_1+B_1+B_3, A_2-C_2+B_2, A_3-C_3	-C_1-D_1+B_3, 0, -C_3	A_1-C_1+B_1+B_3, A_2-C_2, A_3	-C_1+B_3, 0, 0
SM Platform: low-intensity management	A_1+B_1 A_2-C_2+B_2, A_3-C_3	A_1-D_1, 0, -C_3	A_1+B_1, A_2-C_2, A_3	A_1-D_3 A_2, -D_2

3.3. Equilibrium Analysis

Case 1: B_3=0，〖D〗_3=0

(1) From the benefits matrix, in case 1, the platform chooses the strategy "high-intensity management" as strictly inferior to the strategy "low-intensity management", so the social media platform has no incentive to upgrade its technology. When the consumer feels that the extra cost incurred by identifying detrimental information is higher than the loss caused by detrimental information, i.e., C_3＞D_2, an equilibrium strategy combination (low-intensity management, non-self-regulating, non-participatory) is generated. In this case, the platform and information producer can gain benefits, but the information consumer loses benefits due to detrimental information. The network environment also becomes bad and full of all kinds of detrimental information.

(2) A new equilibrium strategy combination (low-intensity management, non-self-regulating, participatory) exists if C_3＜D_2 and the two kinds of benefits from information consumer when the information producer produces high-quality information still can't make up for the extra cost, i.e, A_2+B_2＜C_2.

Case 2：B_3=0，D_3＞0

(1) C_3＞D_2 and when the platform's gain after being punished is still higher than the extra cost when the platform chooses the strategy "low-intensitymanagement", i.e, D_3-A_1＜C_1 the equilibrium strategy combination is (low-intensity management, non-self-regulating, non-participatory). The result is that the network environment does not improve, and the government's measure is ineffective.

(2) D_3-A_1＞C_1，〖A〗_2-C_2＜0, which means the benefit that information producer can gain from the strategy "self-regulating" is not enough to cover the extra cost, a new combination of equilibrium strategy (high-intensity management, non-self-regulating, non-participatory) is generated. In this case, although the detrimental information is successfully deleted, the platform loses its dynamism due to the absence of information on the platform.

(3) In addition there is an equilibrium strategy (low-intensity management, non-self-regulating, participatory) when C_3＜D_2 and A_2+B_2＜C_2.

Case 3：B_3>0，D_3=0

(1) The government subsidy policy is not working when the subsidy cannot cover the additional cost of the social media platform caused by the strategy "high-intensity management", i.e., B_3＜C_1. The result is the same as Case 1.

(2) When B_3＞C_1，〖B〗_3-C_1＜A_1，〖C〗_3＞D_2, the equilibrium strategy combination is (low-intensity management, non-self-regulating, non-participatory), which means the government's subsidy policy is ineffective. If B_3＞C_1，〖B〗_3-C_1＜A_1，〖C〗_3＜D_2，and A_2+B_2＜C_2, there exists an equilibrium strategy combination (low-intensity management, non-self-regulating, participatory), but the government's subsidy policy remains ineffective as before. However, if the earnings from consumed information are more than the extra cost due to the strategy

DOI: 10.54254/2753-7048/37/20240547

"self-regulating", there will be a new strategy combination (high-intensity management, self-regulating, non-participatory). The result is that all three subjects get benefits and detrimental information is effectively managed. It is worth mentioning that in this case, an information consumer can obtain high-quality information even if he does not spend effort to recognize the content of the information.

(3) When B_3-C_1＞A_1, it is not worth to choose the strategy "low-intensity management" for the platform. Therefore, the platform chooses the strategy "high-intensity management" regardless of the strategies chosen by the other two subjects. At this point if A_2＞C_2, we will be able to get the equilibrium strategy combination (high-intensity management, self-regulating, non-participatory).

Case 4: B_3＞0，D_3＞0

(1) When B_3-C_1＜A_1-D_3, the result is the same as in Case 3(2), so no further analysis is needed.

(2) Now discuss the case when A_1-D_3＜B_3-C_1＜A_1. There is an equilibrium strategy combination (high-intensity management, self-regulating, non-participatory) when A_2＞C_2. If A_2＜C_2, there will be a different equilibrium strategy combination (high-intensity management, non-self-regulating, non-participatory). Additionally, when C_3＜D_2， and A_2+B_2＜C_2, we can get the equilibrium strategy combination (low-intensity management, non-self-regulating, participatory).

(3) When B_3-C_1>A_1, the result is the same as in Case 3(3), so no further analysis is needed.

This paper constructs a tripartite game model among a social media platform, an information producer, and an information consumer, and analyzes the equilibrium in four cases according to the change of government policy. The paper summarizes the following conclusions.

First, the government's policy of subsidizing and punishing the platforms can push social media platforms to upgrade their technology in a way. This does help to purify detrimental information at the technical level and create a clean network environment. However, if regulation is excessive and the government forces platforms to control information too tightly it will affect the motivation of network users and make social media platforms lose their advantages.

Second, with or without the government's external intervention, the behavior of information consumers is an important factor in the management of detrimental information. When information consumers believe that the negative impact of detrimental information is bigger than the cost of their time and effort in recognizing the information, they will automatically act as information managers. On the contrary, if information consumers think that detrimental information will not bring about their impact, or even can bring about entertainment effects, they will let the spread of detrimental information. At the same time, when information consumers shift platforms if the loss to platforms and information producers is large enough to break the equilibrium between the three subjects, it will prompt platforms and information producers to make changes.

Third, what information producers choose to publish mainly depends on the benefits they receive from publishing that information. If an information producer can get the attention of Internet users and be able to derive great benefits from publishing detrimental information, but only needs to bear a small portion of the risk, then the information producer will have no incentive to produce high-quality content.

The healthy and orderly development of social media platforms requires cooperation and mutual governance among different network subjects. This paper puts forward the following suggestions.

First, it is necessary to emphasize the important role of information consumers in building a clean and safe network. The government should incorporate quality education on the Internet into the process of compulsory education and social education and raise Internet users' awareness of Internet security and the law. Through such education, Internet users should be encouraged to obtain

DOI: 10.54254/2753-7048/37/20240547

information from diverse sources, break the information cage, and improve their critical thinking skills. To reduce the cost of information identification for Internet users, relevant organizations should also increase the channels through which Internet users can access correct information. In addition, social media platforms should improve the system of reporting and monitoring, so that information consumers can actively participate in the management of detrimental information. However, information consumers should also be prevented from abusing their right to complain.

Second, the government should clarify the legal responsibilities and obligations of information producers. Under the guarantee of information security, the government should further promote the real-name system among information producers. The real-name system raises the risk and cost of posting detrimental information by information producers and can guide information producers to properly utilize their influence. At the same time, the government should adequately protect the intellectual property and legitimate interests of information producers. Relevant departments should help information producers properly deal with malicious reports, plagiarism, and other disputes of interest by actively publicizing related legal knowledge and providing legal assistance, to reduce losses. Platforms should also adjust the incentives for information producers to attract more information producers to produce high-quality content, thereby expanding their influence.

Third, the government should establish a clear system of rewards and punishments for social media platforms by formulating comprehensive laws and regulations related to platform governance. While the government regulates the operation of social media platforms, it should prevent over-regulation and adjust relevant policies timely according to the differences in network environment and social needs. In addition, the government ought to promote healthy competition among social media platforms so that Internet users have enough choice of platforms.

In the end, social media entities and content creators must acknowledge their pivotal role in the spread of information. It is imperative that they exercise social responsibility, cultivate a reputable social presence, and contribute positively to the public discourse. Similarly, consumers of information should enhance their digital literacy, approach social media content with a critical and unbiased perspective and foster independent critical thinking.

4. Conclusion and Recommendations

This paper constructs a tripartite game model among a social media platform, an information producer, and an information consumer, and analyzes the equilibrium in four cases according to the change of government policy. The paper summarizes the following conclusions.

First, the government's policy of subsidizing and punishing the platforms can push social media platforms to upgrade their technology in a way. This does help to purify detrimental information at the technical level and create a clean network environment. However, if regulation is excessive and the government forces platforms to control information too tightly it will affect the motivation of network users and make social media platforms lose their advantages.

Second, with or without the government's external intervention, the behavior of information consumers is an important factor in the management of detrimental information. When information consumers believe that the negative impact of detrimental information is bigger than the cost of their time and effort in recognizing the information, they will automatically act as information managers. On the contrary, if information consumers think that detrimental information will not bring about their impact, or even can bring about entertainment effects, they will let the spread of detrimental information. At the same time, when information consumers shift platforms if the loss to platforms and information producers is large enough to break the equilibrium between the three subjects, it will prompt platforms and information producers to make changes.

Third, what information producers choose to publish mainly depends on the benefits they receive from publishing that information. If an information producer can get the attention of Internet users

DOI: 10.54254/2753-7048/37/20240547

and be able to derive great benefits from publishing detrimental information, but only needs to bear a small portion of the risk, then the information producer will have no incentive to produce high-quality content.

The healthy and orderly development of social media platforms requires cooperation and mutual governance among different network subjects. This paper puts forward the following suggestions.

First, it is necessary to emphasize the important role of information consumers in building a clean and safe network. The government should incorporate quality education on the Internet into the process of compulsory education and social education and raise Internet users' awareness of Internet security and the law. Through such education, Internet users should be encouraged to obtain information from diverse sources, break the information cage, and improve their critical thinking skills. To reduce the cost of information identification for Internet users, relevant organizations should also increase the channels through which Internet users can access correct information. In addition, social media platforms should improve the system of reporting and monitoring, so that information consumers can actively participate in the management of detrimental information. However, information consumers should also be prevented from abusing their right to complain.

Second, the government should clarify the legal responsibilities and obligations of information producers. Under the guarantee of information security, the government should further promote the real-name system among information producers. The real-name system raises the risk and cost of posting detrimental information by information producers and can guide information producers to properly utilize their influence. At the same time, the government should adequately protect the intellectual property and legitimate interests of information producers. Relevant departments should help information producers properly deal with malicious reports, plagiarism, and other disputes of interest by actively publicizing related legal knowledge and providing legal assistance, to reduce losses. Platforms should also adjust the incentives for information producers to attract more information producers to produce high-quality content, thereby expanding their influence.

Third, the government should establish a clear system of rewards and punishments for social media platforms by formulating comprehensive laws and regulations related to platform governance. While the government regulates the operation of social media platforms, it should prevent over-regulation and adjust relevant policies timely according to the differences in network environment and social needs. In addition, the government ought to promote healthy competition among social media platforms so that Internet users have enough choice of platforms.

In the end, social media entities and content creators must acknowledge their pivotal role in the spread of information. It is imperative that they exercise social responsibility, cultivate a reputable social presence, and contribute positively to the public discourse. Similarly, consumers of information should enhance their digital literacy, approach social media content with a critical and unbiased perspective, and foster independent critical thinking.

References

[1] *We Are Social,Meltwater.Digital 2023 July Global Statshot Report[EB/OL].(2023-07-20)[2023-10]. https://wearesocial.com/cn/blog/2023/07/social-media-use-reaches-new-milestone.*

[2] *Liu Jia, Zhang Lin. Research and suggestions on mobile network bad information governance technology [J]. China Informatization,2018(06):61-63.*

[3] *Liang Yanliang, Zhang Mingliang, Shen Yi, Lan Liang. Artificial Intelligence-based Bad Information Supervision System for Internet Data Centers [J]. Network Security Technology & Application, 2022(04):23-25.*

[4] *Pei Zhuoxiong, Yang Min, Yang Jing. Bad information identification technology based on TextCNN-Bert fusion model [J]. Cyber Security And Data Governance, 2023,42(08):72-76.DOI:10.19358/j.issn.2097-1788.2023.08.012.*

[5] *FANG Rui, YU Junyang, DONG Lifeng. Junk text filtering model based on feature matrix construction and BP neural network [J]. Computer Engineering, 2020,46(8):271-276. DOI:10.19678/j.issn.1000-3428.0055414.*

[6] *WU Yifan, ZHU Longjiao, SHI Junping. Application of Artificial Neural Network in Information Filtering [J]. Journal of Jishou University (Natural Sciences Edition) , 2019,40(03):17-22.DOI:10.13438/j.cnki.jdzk.2019.03.004.*

DOI: 10.54254/2753-7048/37/20240547

[7] *U. V G ,V. M M ,D. A A .Correction to: Detection and moderation of detrimental content on social media platforms: current status and future directions[J].Social Network Analysis and Mining,2022,12(1):*

[8] *LIN Ling, CHEN Fuji. The Quadratic Evolutionary Game of Social Media Opinion Dissemination Involving Network Pushers [J]. Journal of Systems Science and Mathematical Sciences, 2023,43(02):379-398.*

[9] *WANG Renda, LI Haifeng. The Technological Innovation of Social Media in the Information Age: Based on the tripartite evolutionary game analysis [J]. Science Technology and Industry, 2023,23(13):139-147.*

[10] *ZHANG Bin, SUI Yu-jia. Strategy Selection of Harmful Information Governance Subjects on Social Media Platforms Based on Perspective of Three - party Evolutionary Game [J]. Journal of Beijing University of Posts and Telecommuni, 2020,22(06):19-29+89.DOI:10.19722/j.cnki.1008-7729.2020.0172.*

[11] *Xiaoyang Sun. Research on Influencing-Subject Game and Control Strategies for Social Media Information Quality [D]. Jiangsu University,2016.*

[12] *Wang Yiming, Xia Zhijie, Luo Mengying. A Game Study on Improving the Quality of User Group Interaction in the Spread of Fake News of Social Media [J]. Journal of Information, 2019,38(12):98-106+140.*

[13] *LUO Meng-ying, XIA Zhi-jie, ZHAI Yue, HE Yin. The Research on the Control of Unconfirmed Information in Social Media Using Game Theory [J]. Information Science, 2017,35(09):44-48.DOI:10.13833/j.cnki.is.2017.09.007.*

[14] *Zhang Yuying. Social media platform governance in multi-party games [D]. Shanghai Academy of Social Sciences,2019.*

[15] *MO Zu-ying. Analysis of Game Theory Model on the Internet Participants Based on Information Quality Control [J]. Information Science, 2013,31(09):137-141.DOI:10.13833/j.cnki.is.2013.09.015.*

[16] *Li Gang, Song Qiang. Game Analysis on China's Internet Vulgar Content Regulation [J]. Management Review, 2011,23(10):77-82.DOI:10.14120/j.cnki.cn11-5057/f.2011.10.001.*

[17] *Smith, Adam. 2012. Wealth of Nations. Wordsworth Classics of World Literature. Ware, England: Wordsworth Editions.*

[18] *Robert J. Aumann.Agreeing to Disagree[J].The Annals of Statistics, Vol. 4, No. 6. (Nov., 1976), pp. 1236-1239*