

Quantitative Analysis of Social Media Influence on Public Opinion

Shuang Wu^{1*}

¹ Nanjing Foreign Language School, Jiangsu, China

Email Address

wushuang901031@163.com (Shuang Wu)

*Correspondence: wushuang901031@163.com

Received: 13 September 2024; Accepted: 30 September 2024; Published: 21 October 2024

Abstract:

This study reveals the mechanisms through which social media shapes public opinion by collecting and analyzing extensive data from various platforms. Our central question focuses on how social media influences the formation of public opinion through information dissemination and user interactions. To address this question, we employ quantitative analysis methods, utilizing theories and techniques from statistics and network science to design a comprehensive analytical framework. This framework encompasses several stages, including data collection, preprocessing, modeling, analysis, and interpretation of results. The goal is to provide a systematic research approach that offers new perspectives and empirical evidence on the impact of social media on public opinion.

Keywords:

Social Media, Public Opinion, Quantitative Analysis

1. Introduction

With the rapid advancement of internet technology, social media has become a crucial platform for people to obtain information, express opinions, and engage in public discussions. In this digital age, social media not only changes the way people live but also profoundly influences the formation and development of public opinion. As a reflection of societal psychology, public opinion plays a significant role in government decision-making, corporate strategy, and individual behavior. Therefore, understanding how social media impacts public opinion is particularly important. Through this research, we aim to uncover the patterns of public opinion formation in the context of social media, providing valuable insights for government agencies, media organizations, and the general public. Additionally, this study will explore how social media platforms can promote a healthy and positive public opinion environment by optimizing algorithms and enhancing transparency.

2. Methodology

2.1. Data Source

Social media platforms, as the primary data source for this study, provide a wealth of user-generated content (UGC). Given the widespread influence of Weibo in China's social media landscape, Weibo was chosen as the main platform for data collection. Utilizing API interfaces and web scraping techniques, we collected Weibo posts related to specific trending events from January 2023 to July 2024. To ensure the representativeness and completeness of the data, we also included user comments and reposts, allowing us to capture the dynamic changes in public opinion comprehensively.

2.2. Analysis Tools and Techniques

Python was employed as the primary programming tool for this study, integrating libraries such as NumPy and Pandas for data processing and analysis. To extract sentiment and thematic content from the text, we utilized natural language processing toolkits like NLTK and TextBlob. Additionally, NetworkX was used to visualize social network structures and dissemination paths. This multi-tool approach facilitates an in-depth analysis of information dissemination characteristics on social media from various perspectives.

2.3. Quantitative Analysis Model

To quantify the impact of social media on public opinion, we developed a network-based propagation model. This model draws from the classic SIR (Susceptible, Infected, Recovered) model, but is extended to suit the characteristics of social media. The model defines three states: Susceptible (S), Active Spreaders (Infected, I), and Non-active Spreaders (Recovering, R). Susceptible users are those who might be influenced by public opinion but have not yet expressed their views [1]; Active Spreaders are users who actively participate in discussions and spread specific opinions; Non-active Spreaders are those who were once active but are no longer engaged in spreading information.

The state transitions in the model can be described by the following differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS \\ \frac{dI}{dt} &= \beta IS - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

Here, β represents the transmission rate, indicating the probability of Susceptible users becoming Active Spreaders; γ represents the recovery rate, indicating the probability of Active Spreaders becoming Non-active Spreaders. These parameters will be determined by fitting the model to actual data, enabling it to better reflect the reality of opinion propagation on social media.

In addition to the propagation model, we employed sentiment analysis algorithms to evaluate the sentiment trends on social media. By scoring the sentiment of textual content, we can quantify the changes in public attitudes toward specific events [3]. Sentiment analysis algorithms help reveal the positive or negative tendencies of public opinion and further explore the relationship between emotional fluctuations and opinion changes.

3. Empirical Analysis

3.1. Data Preprocessing

Data preprocessing is a crucial step to ensure the quality of subsequent analyses. This study first cleansed the raw data by removing irrelevant information, duplicate records, and obvious errors. Then, using regular expressions and natural language processing techniques, we filtered out noise from the text, such as punctuation, stop words, and non-Chinese characters. To further standardize the data, we implemented normalization processes, including uniform date formats, converting text to lowercase, and normalizing numerical data [5].

Additional steps were taken to enhance data quality during preprocessing. The Z-score method and Interquartile Range (IQR) were employed to identify and handle outliers, ensuring the data set did not include points significantly deviating from the normal distribution. For missing data, different strategies were applied based on data type: mean or median imputation for numerical data, and forward or backward filling for categorical data. Regression imputation based on related variables was also considered for a more accurate estimation of missing values [2]. In terms of feature engineering, new features were created to enhance model performance, such as extracting date and hour information from timestamps or constructing composite features based on users' historical behaviors.

3.2. Analysis of Social Media User Behavior

To understand the interaction behavior of social media users and the motivations behind them, various statistical methods were used to analyze user behavior patterns. By calculating metrics such as posting frequency, number of likes, and comment activity, we gained insights into user engagement.

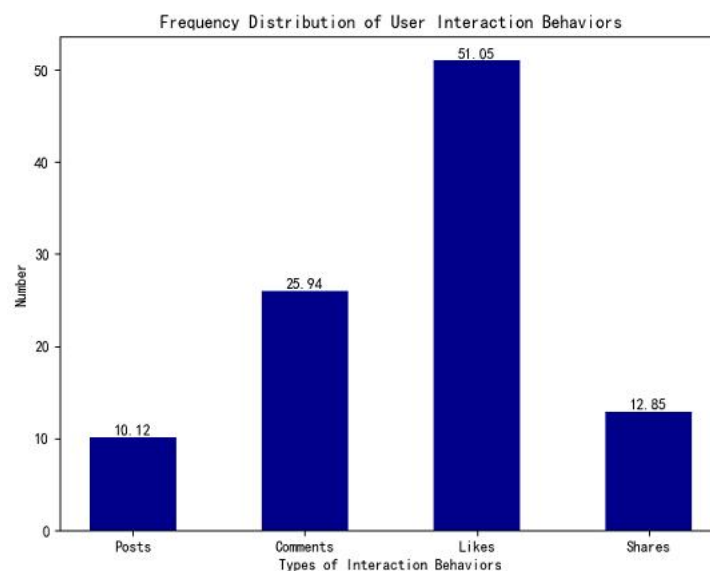


Figure 1. Frequency Distribution of User Interaction Behaviors.

For instance, we observed a significant increase in posting frequency at the onset of certain trending events, indicating heightened public attention. We also noticed that user engagement tends to fluctuate over time, reflecting changes in public interest. Further analysis of user interactions, including mentions, reposting behavior, and frequency of shared topics, revealed the strength of connections among users and the

structural characteristics of the social network. Key users often play critical roles in the network, triggering broader discussions and dissemination. Figure 1 shows the frequency distribution of several typical interaction behaviors of users on a social media platform. The horizontal axis represents several common types of interactions, such as posting, commenting, liking, and sharing, while the vertical axis indicates the average daily number of these interaction behaviors.

3.3. Dynamics of Public Opinion

The evolution of public opinion over time is a major focus of this study. By applying time series analysis methods, we examined how sentiment trends change over time. Specifically, sentiment analysis was applied to textual data at each time point to obtain sentiment scores for public opinion on specific events. These scores were used to construct time series plots, visually depicting the trajectory of sentiment changes.

To quantify these changes, we employed a simple time series model, the moving average model. Assuming Y_t represents the sentiment score at time point t , the moving average model can be expressed as:

$$\hat{Y}_t = \frac{1}{n} \sum_{i=1}^n Y_{t-i+1}$$

where n is the length of the moving window, and $\hat{Y}_t$ is the predicted sentiment score at time point t . This model helps smooth the data and reveal underlying trends.

3.4. Influence Assessment of Key Opinion Leaders (KOLs)

Key Opinion Leaders (KOLs) have significant influence on social media, capable of shaping public opinion considerably [4]. Therefore, quantifying the influence of KOLs is essential. This study utilized centrality metrics in network analysis, including degree centrality, betweenness centrality, and closeness centrality. These metrics help identify users who occupy central positions in the network and are more effective in spreading information.

Degree centrality $C_D(v)$ is the number of edges connected to a node v . It is calculated as:

$$C_D(v) = \sum_{u \in V} a_{uv}$$

where V is the set of all nodes in the network, and a_{uv} is an element of the adjacency matrix, with $a_{uv} = 1$ if there is an edge between nodes u and v , and $a_{uv} = 0$ otherwise.

From Table 1, it can be seen that KOL003 has the highest degree centrality in the network, indicating that this KOL has established the most direct connections with other users. This means that KOL003 is able to disseminate information more effectively and influence a broader user base. The betweenness centrality reflects the importance of a KOL as a pathway for information transmission in the network, and KOL003 also has the highest betweenness centrality, demonstrating its significant role as an intermediary in information dissemination. Closeness centrality measures the average shortest distance from a KOL to other users, and KOL003 has the highest closeness centrality, indicating its superior reachability within the network, enabling it to impact the entire network more quickly.

Table 1. Key Opinion Leader Influence Evaluation.

KOL ID	Degree Centrality	Betweenness Centrality	Closeness Centrality
KOL001	245	0.045	0.98

4. Results and Discussion

4.1. Key Findings

The quantitative analysis of social media's impact on public opinion yielded several key findings. First, user interactions on social media significantly increase during major events, indicating heightened public attention. This growth trend then fluctuates as events unfold, reflecting changes in public interest. Second, our research demonstrates that a small number of key opinion leaders (KOLs) play critical roles in the network. These individuals not only spark widespread discussion but also significantly influence the dissemination paths of information. Third, sentiment analysis reveals notable differences in public attitudes towards various topics, with these attitudes shifting throughout different stages of events and closely correlating with the nature of the content shared on social media.

4.2. Interpretation of Results

To further understand these findings, it is essential to explore the underlying mechanisms. High user engagement at the onset of major events reflects social media users' sensitivity and responsiveness to new developments. This phenomenon can be attributed to the characteristics of social media platforms: immediacy and wide reach enable users to quickly obtain information and express their views on events. Interactions among users also boost engagement, especially when certain events resonate strongly with the public.

The influence of key opinion leaders in the network is undeniable. These users gain high trust due to their expertise in specific fields, higher social status, or extensive social networks. Consequently, their opinions are more readily accepted by other users and can guide the direction of public opinion. Our study shows that KOLs not only serve as bridges in information dissemination but also play vital roles in shaping public discourse. They quickly transmit information to broader audiences, and their viewpoints often become central topics of public discussion.

Sentiment analysis uncovers trends in public attitudes towards different topics over time. These changes reflect fluctuations in public emotions and highlight social media's role in amplifying or suppressing specific sentiments. For instance, positive messages tend to elicit more positive responses, while negative news can exacerbate public negative emotions. This amplification effect is particularly pronounced on social media, where individual emotional reactions are swiftly magnified and transformed into collective emotional experiences.

4.3. Recommendations

Based on the research findings, several recommendations are proposed to address the impact of social media on public opinion. Given the high user engagement at the beginning of events, government and relevant authorities should leverage this period to release authoritative information promptly, guiding the public towards accurate perceptions. Considering the crucial role of KOLs, it is important to establish effective communication channels with these opinion leaders, encouraging them to

play a positive guiding role while minimizing the spread of misleading information. For managing public opinion, a more open and inclusive attitude is recommended, promoting the exchange and collision of diverse viewpoints to foster a healthy public discussion environment. Lastly, for social media platforms, it is advisable to strengthen content regulation, particularly monitoring false information and harmful content, while also enhancing users' information literacy and critical thinking skills to improve their ability to discern the truth.

5. Conclusions

Based on the research findings, several recommendations are proposed to address the impact of social media on public opinion. Given the high user engagement at the beginning of events, government and relevant authorities should leverage this period to release authoritative information promptly, guiding the public towards accurate perceptions. Considering the crucial role of KOLs, it is important to establish effective communication channels with these opinion leaders, encouraging them to play a positive guiding role while minimizing the spread of misleading information. For managing public opinion, a more open and inclusive attitude is recommended, promoting the exchange and collision of diverse viewpoints to foster a healthy public discussion environment. Lastly, for social media platforms, it is advisable to strengthen content regulation, particularly monitoring false information and harmful content, while also enhancing users' information literacy and critical thinking skills to improve their ability to discern the truth.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] Anstead, N.; O'Loughlin, B. Social media analysis and public opinion: The 2010 UK general election. *Journal of computer-mediated communication*, 2015, 20(2), 204-220.
- [2] Huang J.; Cook, G.G.; Xie Y. Large-scale quantitative evidence of media impact on public opinion toward China. *Humanities and Social Sciences Communications*, 2021, 8(1), 1-8.
- [3] Whitman, M.S. Media Influence on the Public Opinion of Criminal Justice Issues and Policies: A Quantitative Descriptive Study. Northcentral University, 2019.
- [4] Pinto, S.; Albanese F.; Dorso C O, et al. A quantitative analysis of Media Agenda and Public Opinion using time-evolving topic distribution. arXiv preprint arXiv:1807.05184, 2018.
- [5] Akpan, C.S.; Onyebuchi, C.A. Quantitative and qualitative analysis of audience perception of the effect of public opinion polls in Nigeria's nascent democracy. *International Journal of Humanities and Social Science*, 2012, 2(15), 231-242.



© 2024 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)