Postprint: The Interactive Influence Mechanism of Social Bots' Emotions in Online Emergencies

Authors: Ma Xiaoyue, Meng Xiao, Wang Zhen, Liu Yidong

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] To refine the intervention mechanism of social bots on netizens' emotions, and to provide practical suggestions for public opinion governance in online emergencies.

[Method/Process] Taking the Zhao Xiaojing incident at Renji Hospital as a case study, we employed the Naive Bayes method to calculate sentiment orientation on Weibo, and constructed a Vector Autoregression (VAR) model to conduct Granger causality tests, impulse response analysis, and variance decomposition analysis, thereby determining the emotional relationships among social bots, opinion leaders, and ordinary users at various stages of the event life cycle.

[Results/Conclusions] The emotional relationships among social bots, opinion leaders, and ordinary users evolve with the stages of public opinion development. During the outbreak period, social bots amplified the emotional influence of opinion leaders on ordinary users; during the maturity period, the influence of social bots waned, and ordinary users'emotions reacted back upon social bots and opinion leaders; during the decline period, the three parties maintained relatively independent emotional relationships. Furthermore, the influence strategies of social bots exhibit characteristics of concealment and indirectness.

Full Text

Interactive Sentimental Influence Mechanism of Social Bots During Online Emergency Events

Ma Xiaoyue[1]**, Meng Xiao**[1]**, Wang Zhen**[1]**, Liu Yidong**[2] [1]School of Journalism and New Media, Xi' an Jiaotong University, Xi' an 710049 [2]Shanghai Pudong Weirendian Big Data Research Institute, Shanghai 201203

Abstract: [Purpose/Significance] This study elaborates on the intervention mechanism of social bots on netizens' sentiment and provides practical recommendations for online emergency event governance. [Method/Process] Taking the Zhao Xiaojing incident at Renji Hospital as a case study, we calculated Weibo sentiment orientation using the Naive Bayes method, and determined the emotional relationships among social bots, opinion leaders, and ordinary users at various stages of the event lifecycle by constructing a Vector Autoregression (VAR) model, conducting Granger causality tests, impulse response analysis, and variance decomposition analysis. [Result/Conclusion] The emotional relationships among social bots, opinion leaders, and ordinary users evolved with the public opinion stages. During the outbreak stage, social bots amplified the emotional influence of opinion leaders on ordinary users; during the mature stage, the influence of social bots declined, and ordinary users' sentiment reacted upon social bots and opinion leaders; during the decline stage, the three parties maintained relatively independent emotional relationships. Additionally, the influence strategies of social bots exhibited insidious and indirect characteristics.

Keywords: emergency event; social bots; sentiment analysis; time series analysis

1 Introduction

The proliferation of social media in recent years has provided rapid diffusion channels for emergency events. Increasing research has found that information published by social bots constitutes a substantial proportion of social media content [1]. The online participation of social bots can easily interfere with public cognition and judgment of events, foster negative emotions among netizen groups, and enhance the emotional polarity of online discussions [2], significantly increasing the uncertainty of online emergency event public opinion. Therefore, examining the interactive emotional influence between social bots and human users from a sentiment perspective during the evolution of online emergency events holds important practical governance significance.

Most existing research on social bots is based on foreign social media platforms such as Twitter, lacking studies in the Chinese social media environment. Secondly, sentiment is the fundamental force and deep logic driving the generation, formation, and evolution of online public opinion [6]. The interactive characteristics between social bots and human emotions reflect the potential core of human-machine relationships. Previous studies have emphasized static descriptions of social bot content and behavioral features, with less verification of the statistical significance of interactive influences between social bots and human emotions. Moreover, traditional research on online emergency events has primarily focused on government departments, online media, ordinary netizens, and paid posters as public opinion subjects, with insufficient attention to social bots.

In light of this, this study attempts to embed social bots as a new participant into the emergency event research framework, revealing the emotional interaction relationships among social bots, opinion leaders, and ordinary users in the context of online emergency events, particularly the differences in these relationships across various stages of the event lifecycle. This research addresses the following questions: What emotional relationships exist among social bots, opinion leaders, and ordinary users, and how do these relationships change across different stages of online emergency events? What behavioral strategies do social bots employ to influence human emotions?

To address these questions, this study analyzes actual Weibo data from emergency events, extracting sentiment orientation values for social bots, opinion leaders, and ordinary netizens based on user classification and event lifecycle division, and employs Vector Autoregression models and generalized impulse response functions to analyze the emotional relationships among these three types of participants.

2 Literature Review

2.1 Social Bots Social bots are computer algorithms that automatically generate content and interact with humans on social media, attempting to mimic and potentially alter human information behavior [7]. Some social bots provide services, such as automatic alarms during natural disasters and crises [8], while others are used as "online water armies" with social harmfulness. Existing research on social bots primarily focuses on two aspects: social bot identification and monitoring; social bot activities and their impacts.

In terms of social bot identification, methods relying on social network characteristics to distinguish real users from social bots have been widely used. Y. Boshmaf et al. proposed an automatic fake account detection system called Integro, which uses supervised machine learning and user ranking schemes to rank most real accounts higher than fake accounts [9]. S. Hurtado et al. assumed that accounts with highly correlated temporal activity are likely bots, using temporal and network information to detect political bots on the Reddit platform [10]. Another common identification method is machine learning-based; the famous BotOrNot system trains classifiers using over a thousand behavioral features to measure the likelihood of an account being a social bot [11]. In addition to behavior, text information and content analysis have also been attempted for social bot identification [12].

Regarding social bot behavioral characteristics and impacts, researchers have extensively focused on social bots' performance in political activities. Social bots are generally considered to manipulate social media public opinion, change public perception of political entities, and even influence political election outcomes [13]. For example, during the 2016 U.S. election, some candidates used automated accounts or bots to increase their social media attention and follower counts [14]. In Brexit, numerous social bots only posted tweets during the referendum debate and disappeared after voting [15]. Regarding Chinarelated topics on Twitter, research suggests that social bots can increase human users' exposure to specific information and alter existing information interaction structures [16]. Zhang Hongzhong summarized five strategies of political bots: creating false popularity, pushing large amounts of political messages, spreading false or spam political information, creating smoke screen effects to confuse the public, and shaping highly personalized virtual opinion leaders [17]. However, some studies suggest that social bots' influence is not significant; F. Brachten et al. found that during the 2017 German state elections, identified social bots showed no signs of collective political strategies and had minimal influence [18]. In addition to political topics, social bot participation has also been found in public health issues such as anti-vaccination movements [19] and e-cigarette promotion [20].

2.2 Sentiment Analysis in Online Emergency Events Sentiment analysis involves analyzing people' s sentiments, attitudes, and opinions toward different entities. Sentiment analysis in online emergency events is important for understanding netizen viewpoints and decision-making, with research primarily focusing on information monitoring, sentiment evolution patterns, emotion diffusion patterns, and government interaction strategies. At the information monitoring level, scholars have established emergency event themes and sentiment monitoring models to reflect dynamic sentiment transfers across regions [21]. Regarding sentiment evolution, Mao Taitian et al., through analysis of Weibo posts about the Shanghai police tripping a child incident, found that under unclear information conditions, the public is easily influenced by the most prominent opinions in the public opinion field [22]. From an emotion diffusion perspective, S. Stieglitz et al., based on a study of over 165,000 tweets, showed that emotional tweets are more likely to be retweeted than neutral tweets [23]. Regarding the relationship between government information behavior and public emotion, W. Zhang et al.' s research found that government information release strategies have a significant negative impact on the spread of anxiety and disgust emotions [24].

Previous researchers' methodologies mainly include sentiment dictionary-based sentiment computation and machine learning classification algorithms based on neural networks and deep learning [25]. The former primarily determines word sentiment orientation through sentiment annotation of opinion words, with three main methods for establishing sentiment dictionaries: manual, corpus-based, and dictionary-based approaches. The manual method is more time-consuming than the latter two and is usually combined with them to avoid errors from automated techniques [26]. F. Aslam et al. used the NRC sentiment dictionary to calculate the emotional performance of news headlines during the COVID-19 pandemic across eight dimensions [27]. Domestic scholars An Lu et al., taking the Zika event as an example, referenced the Dalian University of Technology sentiment dictionary and used pointwise mutual information methods to build an emoji dictionary for calculating Weibo text sentiment [28]. Zhang Peng et al. combined grounded theory characteristics to construct an emergency event-specific sentiment dictionary, improving sentiment analysis accuracy to some extent [29]. Machine learning methods use a portion of complete data to train classifiers without relying on prior vocabulary, with typical methods including Naive Bayes, Maximum Entropy, and Support Vector Machines. For example, Deng Jun et al. constructed a sentiment classification model using Word2Vec and SVM methods to classify sentiment polarity in comments about the Didi Wenzhou girl murder case [30]. Wu Peng et al. used sentimentally meaningful word vectors as text features and a bidirectional long short-term memory model as a classifier to classify netizen sentiment polarity [31]. Although machine learning methods are widely applied, they require large amounts of training data and high-quality training data, and also suffer from black box and interpretability issues.

In summary, previous research has made numerous attempts in social bot identification and sentiment analysis techniques, but several shortcomings remain: Although prior research on social bot impacts has emphasized their manipulation of online public opinion, this influence has rarely been verified, with more studies focusing on describing social bot content and behavioral features, and cases mostly from foreign social media and political topics. Existing sentiment analysis in online emergency events has primarily involved government, media, and opinion leaders, lacking attention to social bots, and no research has placed the sentiment of social bots, opinion leaders, and ordinary users as three variables in the same time series system for analysis. Therefore, it is necessary to test the emotional relationships among social bots, opinion leaders, and ordinary users in the context of online emergency events to provide references for public opinion monitoring and management.

---

3 Research Methods

This study takes a specific online emergency event as a case, using Vector Autoregression models to analyze the interactive emotional influence among social bots, opinion leaders, and ordinary users. The research process can be divided into three parts: data acquisition and preprocessing, sentiment orientation extraction, and emotional relationship analysis. First, we preprocess the acquired social bot and human Weibo data, removing meaningless information and dividing participants and event lifecycle stages. Then, we calculate Weibo sentiment orientation to obtain emotional change curves for social bots, opinion leaders, and ordinary users. Finally, we establish VAR models for the emotional time series of social bots, opinion leaders, and ordinary users at each event stage to analyze the emotional interactions among the three types of participants.

3.1 Data Acquisition and Preprocessing

3.1.1 Data Acquisition and Text Cleaning This study selected the "Zhao Xiaojing incident at Renji Hospital" as the emergency event case. The incident occurred on April 24, 2019, when Shanghai Renji Hospital expert Zhao Xiaojing was handcuffed and taken away by police for refusing to treat a patient who had cut in line and arguing with the patient. The incident was exposed on the Sina Weibo platform on April 26, 2019, triggering widespread social attention. The Weibo dataset used in this study was provided by the Sina Weirendian Big Data Research Institute, a big data communication application platform under Sina Weibo that provides new media big data services. The dataset was desensitized, containing 351,257 Weibo posts related to the incident before and after its occurrence, with annotations indicating whether the publisher was a social bot. Each data entry included retweeted Weibo content, Weibo posting time, original Weibo content, original Weibo posting time, verification type, and social bot identity annotation. Due to intellectual property protection issues, this dataset is not yet publicly available.

After obtaining the initial Weibo data, we selected relevant Weibo posts from the main discussion period of the event (April 26, 2019 to May 2, 2019), totaling 346,447 posts. We used regular expressions to remove fields that interfere with sentiment analysis results from Weibo text, such as "@username" and "retweeted Weibo."

3.1.2 Opinion Leader and Ordinary User Classification In the two-step flow theory, opinion leaders are important intermediaries in information dissemination [32]. They often have high social status and extensive social networks, and can influence and shape others' viewpoints [33]. To distinguish opinion leaders from netizens and other subjects, this study subdivides human users into two categories: opinion leaders and ordinary users. Verified users on Weibo are often experts and celebrities from various industries, and verification status plays an important role in opinion leader influence [34]. To extract opinion leaders from the event, this study classifies human users based on verification type, categorizing users with verification types of "Gold V," "Orange V," "Blue V," and "Expert" as opinion leaders, and other human users as ordinary users.

3.1.3 Lifecycle Division Scholars have divided the lifecycle of online public opinion differently according to their research purposes. For example, Wang Guohua et al. divided the public opinion of the Yao Jiaxin incident into four stages: occurrence, development, evolution, and dissipation [35]; Li Gang et al. divided online public opinion of emergency events into four stages: latent, outbreak, mature, and decline [36]; Xie Kefan et al. divided online public opinion emergency events into five stages: incubation, germination, acceleration, maturity, and decline [37].

Combining previous division methods and considering the short lifecycle and rapid development of the selected event' s public opinion, this study divides the "Zhao Xiaojing incident at Renji Hospital" into three main stages—outbreak, maturity, and decline—based on the Weibo discussion volume curve.

3.2 Sentiment Orientation Analysis of Social Bots, Opinion Leaders, and Ordinary Users SnowNLP is a Chinese natural language processing library inspired by TextBlob, providing a text sentiment orientation calculation method based on the Naive Bayes algorithm. Since the initial sentiment analysis function used shopping reviews as training data, this study trained a sentiment orientation discrimination model using manually sentiment-annotated Weibo corpora to make it suitable for Weibo sentiment analysis tasks, achieving 90.5% sentiment classification accuracy on the test dataset. The model was then used to determine sentiment orientation, obtaining the probability of a single Weibo text being positive. Probability values range between 0 and 1, with values closer to 1 indicating more positive text sentiment and values closer to 0 indicating more negative sentiment. Considering the brevity of emergency event public opinion cycles, we segmented the Weibo data into 10-minute slices, calculated the average sentiment value for each time slice, and ultimately obtained time series data and sentiment evolution curves for opinion leaders, social bots, and ordinary users.

3.3 Emotional Relationship Analysis Among Social Bots, Opinion Leaders, and Ordinary Users The sentiment of social bots, opinion leaders, and ordinary users exists in a dynamically interactive system. Traditional multiple linear regression struggles to explain this, so this study employs the Vector Autoregression model (VAR) commonly used in economics to analyze the emotional relationships among social bots, opinion leaders, and ordinary users. We established separate VAR models for the three main stages of public opinion (outbreak, maturity, decline), and conducted Granger causality analysis, generalized impulse response function analysis, and variance decomposition analysis based on the VAR models.

3.3.1 VAR Model Construction When time series data is unstable, it may lead to spurious regression and affect result reliability. Therefore, before using the VAR model, we first conducted ADF unit root tests on the data from each lifecycle stage. All P-values were significant, indicating that each group of variable data was stable and could directly undergo Granger causality tests and VAR model construction.

VAR model construction requires determining the optimal lag order to balance lag period and degrees of freedom. This study comprehensively considered LR, FPE, AIC, SC, and HQ test criteria to determine that the optimal lag orders for the outbreak, maturity, and decline stage models were lag 1, lag 1, and lag 2, respectively. We then established VAR models based on the optimal lag periods. The corresponding VAR(p) model form is as follows:

$$y_t = A_0 + A_{1y_{t-1}} + \cdots + A_{py_{t-p}} + e_t$$

where $y_t = \begin{bmatrix} \text{OpinionLeader}_t \\ \text{Crowd}_t \\ \text{SocialBots}_t \end{bmatrix}$ represent the sentiment orientations of opinion leaders, ordinary users, and social bots, respectively; A_0 is the constant vector; e_t is the random disturbance term matrix; A_t is the coefficient matrix; and p is the lag order.

To determine model stability and result correctness, we conducted robustness tests on the models. AR root test results showed that the moduli of characteristic roots were all less than 1, indicating model stability and no need for reconstruction.

3.3.2 Granger Causality Test To determine the influencing and influenced relationships among the sentiments of social bots, opinion leaders, and ordinary users, we applied Granger causality tests. The basic idea of Granger causality testing is that if one time series variable x is a Granger cause of another variable y, it means that predictions of variable x using a model that includes y' s past values are more accurate than predictions using only x' s past values. In the VAR(p) model of this study, to test whether subject A' s sentiment y_j is a Granger cause of subject B' s sentiment y_i, the null hypothesis of the Granger causality test is that subject A' s sentiment is not a Granger cause of subject B' s sentiment, i.e., the parameters $a_{ij}^{(q)}$, $q = 1, 2, 3, \dots, p$ of y_j are collectively 0. Granger test results are completed through F-tests.

3.3.3 Generalized Impulse Response Function Analysis Impulse response functions describe the impact of a random error term shock from one endogenous variable on another endogenous variable. The basic idea of impulse response functions, illustrated with a VAR(2) model, is as follows:

$$x_t = a_{1x_{t-1}} + a_{2x_{t-2}} + b_{1z_{t-1}} + b_{2z_{t-2}} + \varepsilon_{1t}$$
$$z_t = c_{1x_{t-1}} + c_{2x_{t-2}} + d_{1z_{t-1}} + d_{2z_{t-2}} + \varepsilon_{2t}$$

Assuming the above system starts activity at $t = 0$, with $x_1 = x_2 = z_1 = z_2 = 0$, and setting disturbance terms $\varepsilon_{10} = 1$, $\varepsilon_{20} = 0$ in period 0, with all subsequent terms being 0. The disturbance given initially will continuously transmit through the system. Through iterative calculation, we can obtain $x_0, x_1, x_2, x_3, \dots$, called the response function of x caused by x' s impulse. Similarly, we can obtain $z_0, z_1, z_2, z_3, \dots$, called the response function of z caused by x' s impulse. When the impulse in period 0 is $\varepsilon_{10} = 0$, $\varepsilon_{20} = 1$, we can obtain the response functions of x and z caused by z' s impulse. This study uses the generalized impulse response function method proposed by H. H. Pesaran and Y. Shin for analysis, whose results are not affected by variable ordering relationships [38]. In this study, generalized impulse response function analysis results reflect the magnitude, direction, and lag characteristics of emotional changes in one type of subject (opinion leaders, ordinary users, or social bots) caused by emotional fluctuations in another type of subject.

3.3.4 Variance Decomposition Analysis Variance decomposition can characterize the contribution degree of structural shocks to the variation level of the studied endogenous variables, thereby evaluating the importance of different structural shocks. The Vector Moving Average Model (VMA) obtained from equation (1) is:

$$y_{it} = \sum_{j=1}^{k} \left(c_{ij}^{(0)} \varepsilon_{jt} + c_{ij}^{(1)} \varepsilon_{jt-1} + c_{ij}^{(2)} \varepsilon_{jt-2} + \ldots \right)$$

where $c_{ij}^{(q)}$ is the response function of y_i caused by y_j' s impulse in period q. Accordingly, the variance decomposition model used in this study is:

$$RVC_{j \to i}(\infty) = \frac{\sum_{q=0}^{\infty} \left(c_{ij}^{(q)} \right)^2 \sigma_{jj}}{VAR(y_{it})} = \frac{\sum_{q=0}^{\infty} \left(c_{ij}^{(q)} \right)^2 \sigma_{jj}}{\sum_{j=1}^{k} \left\{ \sum_{q=0}^{\infty} \left(c_{ij}^{(q)} \right)^2 \sigma_{jj} \right\}}$$

$RVC_{j \to i}(\infty)$ is the relative variance contribution, ranging between 0 and 1, observing the influence of variable j on variable i based on the relative contribution of variable j' s shock variance to y_i' s variance. $\sum_{q=0}^{\infty} \left(c_{ij}^{(q)} \right)^2 \sigma_{jj}$ is the result of evaluating the impact of the jth disturbance term on the ith variable from infinite past to present using variance. In practice, if the model satisfies stationarity conditions, $c_{ij}^{(q)}$ decays geometrically as q increases, so only a finite s term is needed [39]. In this study, we use variance decomposition to identify the contribution degree of influence among opinion leaders, ordinary users, and social bots.

---

4 Research Results

4.1 Data Description This study takes the "Zhao Xiaojing incident at Renji Hospital" as an example, using the method described in section 3.2 to plot the Weibo public opinion evolution trend and divide the event' s public opinion lifecycle into three stages: outbreak, maturity, and decline, as shown in [Figure 1: see original paper]. The Weibo dissemination of the Zhao Xiaojing incident at Renji Hospital exhibited a "single-peak" characteristic

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.