A Study on the Propagation of Online Public Opinion by Internet Water Army

Fuzhong Nian (gdnfz@lut.edu.cn)
Lanzhou University of Technology
Chongpei Wang
Lanzhou University of Technology
Duan Zhang
Lanzhou University of Technology
Zhongkai Dang

Research Article

Keywords: Internet Water Army, Online Public Opinion, Information dissemination Model, Online Social Network.

Posted Date: October 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3447339/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Social Network Analysis and Mining on January 17th, 2024. See the published version at https://doi.org/10.1007/s13278-023-01182-w.

A Study on the Propagation of Online Public Opinion by Internet Water Army

First Fuzhong Nian[1*], Second Chongpei Wang[2†], Third Duan Zhang[3†], Fourth Zhongkai Dang[4†]

[1,2,3*]Lanzhou University of Technology, School of Computer & Communication, Lanzhou, 730050, China.

[4]The National Computer Network Emergency Response Technical Team & Communication Engineering, Beijing, 100083, China.

*Corresponding author(s). E-mail(s): gdnfz@lut.edu.cn;
Contributing authors: 2807387609@qq.com; wangpiang12650@outlook.com; 18810520671@163.com;
†These authors contributed equally to this work.

Abstract

In the dissemination of information on the Internet, internet water armies have gradually become an important influencing factor in the direction of public opinion. This paper takes Weibo as the research background, obtains real data, establishes a communication dynamics model, and quantitatively researches the influence of internet water armies on the dissemination of public opinion in social networks. Firstly, this paper obtains real hot events in Weibo, analyzes the number of retweets and comments under the same hot topics, and finds the users of the internet water armies. Secondly, this paper establishes a new social network SI_wI_cS propagation model based on the propagation mechanism among users in Weibo, combining with the SIS infectious disease model, and introducing the impulse effect, the heat effect, and the herd effect, to realize the research of public opinion propagation in Weibo. After that, the SI_wI_cS propagation model is applied to small-world and scale-free networks to simulate the propagation of public opinion in social networks and analyze the influence of internet water armies on the propagation. The experimental results summarize the influence of network water armies on the propagation of public opinion. Finally, the acquisition of real data is utilized to verify the rationality and validity of the model. This paper quantitatively considers the driving nature of internet water armies on the trend of public opinion in the communication dynamics model, which provides a reasonable decision-making basis for the dissemination and control of public opinion in social networks, and has important theoretical and practical significance.

Keywords: Internet Water Army, Online Public Opinion, Information dissemination Model, Online Social Network.

1 Introduction

In recent years, with the emergence of various social apps, more and more Internet users have begun to use Weibo, headlines and jitterbugs to obtain information about hot social events. At the same time, a group of hired Internet writers who publish specific information for specific content in the network have also emerged, and they are the Internet water armies. Network water armies are usually active in social networking platforms such as forums, Weibo, video sharing sites, etc. , and by disguising themselves as ordinary netizens or consumers, they influence normal users by posting, recovering and spreading blog posts (Jiaze (2021); Liu and Xia (2021); SHEN et al (2017);

Shin et al (2011); Chen et al (2016); Geng et al (2023); Wu et al (2023)). Taking advantage of the fact that netizens do not know the facts, they cause adverse effects by guiding the direction of online public opinion and disrupting the order of the network (Guo et al (2023); Xu et al (2013); Jain et al (2023); Wang and Li (2023)). With the emergence of Internet water armies, the concocted other-generated public opinion has overshadowed the naturally formed social opinion, which not only harms the healthy expression of public opinion on the Internet, but also causes the distortion of the data of the Internet public opinion monitoring and statistics, which is misleading to the academic research institutes and the government, as well as puts some social contradictions at the risk of being intensified or amplified. The identification of Internet water armies on the Internet and their influence on the dissemination of Internet public opinion have received increasing attention from researchers in different fields, including computer science, economics, mathematics, finance, etc. (Chen et al (2013); Zeng et al (2014); Wang et al (2014)).At present, many foreign scholars study the identification of Internet water armies for false comments (Jabeur et al (2023); Bathla et al (2022); Vidanagama et al (2022); He et al (2022); Lee et al (2022)); domestic scholars mainly study the identification of Internet water armies and the governance of rumors (He et al (2023); Li et al (2014); Peng et al (2023); Zhang et al (2019); Yan et al (2023); Chen and Du (2023); Zhang et al (2023)). However, most of these researches use different algorithms to accurately identify Internet water armies, and there is a lack of research on how Internet water armies promote the development of Internet public opinion communication.

On the basis of the previous research, this paper conducts a systematic analysis and research on how Internet water armies carry out diffusion and dissemination of Internet public opinion. In order to better understand the diffusion behavior of network water armies on network public opinion and the influencing factors involved in the diffusion. A large number of Weibo data sets were collected at the beginning of the study, and these data were counted and observed. The number of retweets and comments on a single Weibo sent by a user under the event of a hot Weibo was counted (Chen et al (2015)), as well as for the number of retweets of Weibo conforming to the way of using the topic symbol "#" and external link URLs by the users of the water armies (Bindu et al (2018)). Drawing on Wang et al.(2022) and Angel Tocino et al.(2023) on the SIS epidemic transmission model, a new $\mathrm{SI_wI_cS}$ network water army transmission model is constructed by improving on the original SIS transmission model. The infector I is divided into two states, called I_w and I_c. where I_w is named as the original Internet water army infector, which serves as a set part of the Internet water army at the beginning of the propagation. I_c is called the common infector, which is transformed from the original Internet water army by infecting the susceptible person S. It maintains the infected state of the consistent viewpoints of the topic with the original water army, and at the same time, has the ability of infecting the susceptible person S. I_c will change back to the susceptible S state with a certain probability under different external effects, but I_w will not change and will always maintain a single-direction propagation state. Drawing on the definition of Internet water armies by previous scholars, and then according to the data set obtained from Weibo screening, we get the trend of Weibo's Internet water armies' propagation of network public opinion and carry out the corresponding research, and we can accurately predict the propagation trend of Internet water armies' propagation of network public opinion through the model of $\mathrm{SI_wI_cS}$ Internet water armies' propagation.

2 Model profile

Firstly, a susceptible-infected-susceptible $\mathrm{SI_wI_cS}$ Internet water army propagation model is established, assuming that the propagation of Internet public opinion is spread through the infected nodes of the Internet water army, and now the population is divided into susceptible population S and infected population I. Where S and I represent the susceptible population S who have never heard of the public opinion, and the water army population I who are spreading the public opinion, respectively. where I contains two state mechanisms I_w and I_c. I_w as the original Internet water army, will not change its state and number, but I_w infects the susceptible people S to make them change to the state of I_c and the viewpoints of I_c are consistent with those of I_w, so as to play the role of guiding the network public opinion and promote the dissemination of network public opinion, and I_c, as a common infected person, has the main role of promoting the network public opinion. I_c as a common infected person has a major role in promoting online public opinion. From now on, we will refer to the original Internet water armies as the "water army", and the ordinary infected people as the "infected people" (the susceptible people who maintain the same opinion direction with the water army and are infected by the water army).

2.1 Schematic of the propagation of nodes

The nodes are infected with different probabilities during the propagation state of the nodes to the susceptible nodes. The initial proceeding is to infect the susceptible person with probability β through the water army I_w, at which time the susceptible node will be infected with I_w as I_c. In the initial stage of public opinion occurrence, the effective transmission rate of both the waterborne army and the infected person for the susceptible person is β. The propagation process is shown in Fig.1. From Fig.1(a) and Fig.1(b), we can see that when the public opinion first starts to spread,

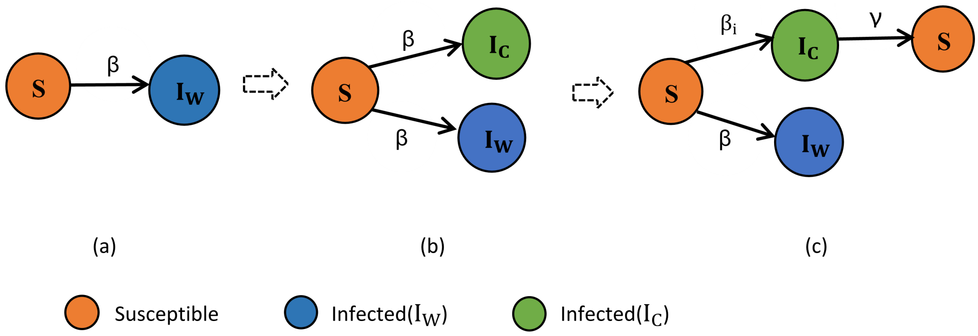


Fig. 1 Schematic diagram of node state infection

the susceptible nodes are infected by the water army and become infected, and then the water army and the infected people infect the susceptible people together, and at this time the camp for the susceptible people to be infected is divided into the water army and the infected people. Fig.1(c) indicates that as time t increases, the effective transmission rate β of the infected person is changed to β_i by the factors of three external effects (impulse effect, heat effect and herding effect). Finally, as time passes, the infected person will become susceptible again with a certain probability γ as the interest in that opinion topic disappears. The details are as follows.
(1)When a susceptible node encounters a sailor node, it will be infected by the sailor node with a *beta* probability to become an infected node.

$$S_i + I_{w_j} \xrightarrow{\beta} I_{w_i} + I_{c_j} \tag{1}$$

(2)A susceptible node encounters an infected node and becomes infected by the infected node with probability β_i and the susceptible node becomes a new infected node.

$$S_i + I_{c_j} \xrightarrow{\beta_i} I_{c_i} + I_{c_j} \tag{2}$$

(3)After an infected node has been affected by a variety of external factors, there is a certain probability *gamma* that it will change to a susceptible state again.

$$I_{c_i} + S_j \xrightarrow{\gamma} S_i + S_j \tag{3}$$

Set the infection rate of the water army node I_w for neighboring nodes to 100% and the infection rate of the infector I_c for neighboring nodes to θ. The infector node does not infect all the neighboring nodes, it is selectively infected with probability of θ. The propagation schematic is shown in Fig.2.

2.2 Dissemination of influencing factors

Considering that in real life, the process of spreading public opinion in the network may be affected by different external effects on the process of spreading. Therefore, three different effects, namely, the crowd impulse effect, the public opinion heat effect, and the crowd crowd effect, are added to the model to simulate different external effects, so as to make the influence of Internet water armies on the dissemination of Internet public opinion more closely match the dissemination of Internet public opinion in reality.

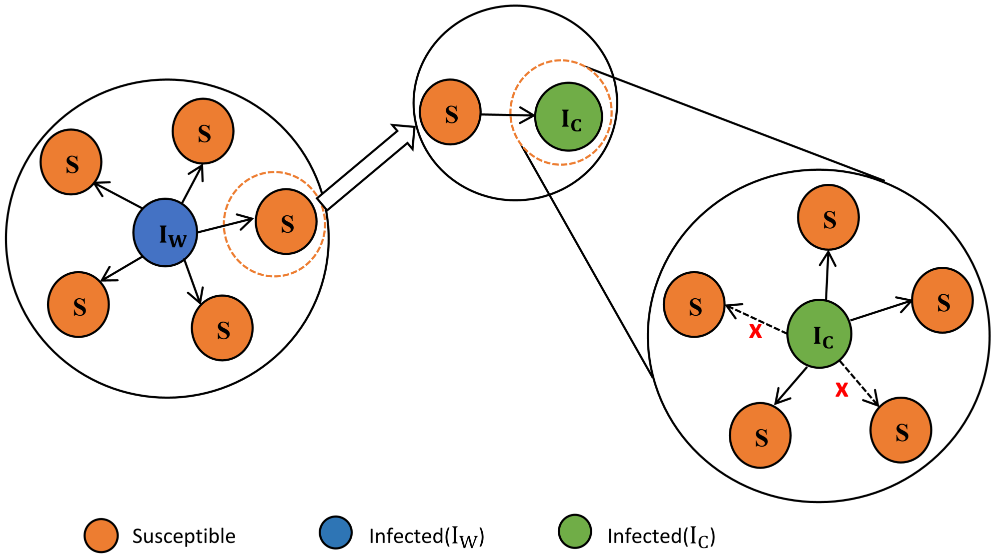


Fig. 2 Infection of neighboring nodes by sailors and infected people

The dynamic propagation probability $\beta_{(i,t)}$ of the infected person under the SI_wI_cS model, where β is the initial effective propagation probability, and these three functions are the impulse effect function $\Gamma_{(t)}$, the heat effect function $\Omega_{(t)}$, and the herding effect function $\Pi_{(t)}$, respectively, with the following equations.

$$\beta_{(i,t)} = \beta \left(\Gamma_{(t)} + \Omega_{(t)} + \Pi_{(t)} \right) \tag{4}$$

$$\beta_{(i)} = \beta \left(\Gamma_{(t)} + \Omega_{(t)} + \Pi_{(t)} \right) \tag{5}$$

According to Nian et al.(2019) concluded that in social networks, information spreads through the network and people's interests change over time. When some information of public opinion first appears in the network, the initiators often have an impulsive mindset and they will spread the information around in the hope that this information can be diffused in the network, which is called the impulsive effect. This effect fades out with time and is defined in the propagation rate as Eq(6). where μ is a constant.

$$\Gamma_{(t)} = e^{\frac{\mu}{t+1}} \tag{6}$$

When a new public opinion starts to spread in the network, people will take great interest in it and the probability of infection will gradually grow. With the passage of time, the news is already well known to people and the interest level is slowly fading, the spread of online public opinion is heat-limited and conforms to the trend of Ebbinghaus forgetting curve (Ebbinghaus (2013)) after a certain growth. This is referred to as the heat effect and defined as Eq(7). where X and δ are constants.

$$\Omega_{(t)} = e^{\frac{(t-X)^2}{\delta}}. \tag{7}$$

The herding effect of people needs to be taken into account during the spread of the water army, and in this regard WAN et al.(2016) and Wang et al.(2017) have already mentioned about the impact that rumors have in social networks in the herding effect. In real life, the herd effect, also known as the band float effect, means that when individuals are influenced (guided or pressured by the group), they will doubt and change their opinions, judgments and behaviors in the direction of being in agreement with the majority of the group. Calling this the herd effect or herding effect, it is defined as Eq(8). where p is a constant that represents the sum of the number of water army and infected people.

$$\Pi_{(t)} = \log_{10}(\frac{I_{(t)} + p}{e^2}). \tag{8}$$

In these 3 dynamic communication effects for the network of water army in the network of communication presents a linear correlation. The stronger the effect of the spread at the beginning, the stronger the spread of the internet water army, and the more infected people will be. It has a greater role in promoting online public opinion. However, as time slowly increases, the strength of various effects will weaken, resulting in weaker dissemination of online public opinion by internet water army, while the role of promoting online public opinion will become smaller.

2.3 $\mathbf{SI_wI_cS}$ model equations

Set the state of the water army to be a constant state, where the number and the probability of propagation do not change. What has changed is the number of infected people who have been infected by the water army. The probability of infection of the susceptible population by infected people changes over time, and the number of infected nodes keeps changing. Consider a closed uniform mixture of N individuals as a social network, where individuals are nodes and contacts between individuals are edges. Under the assumption of uniform mixing, the following parameters are discussed in the uniform network species. $S_{(t)}$, $I_{w(t)}$ and $I_{c(t)}$ what they represent is the population density of susceptible nodes, water army nodes, and infected nodes at moment t, respectively, where k represents the average degree of the network. Considering the propagation mechanism of the water army, the equation of the internet water army model of SI_wI_cS can be expressed as.

$$\begin{cases} \frac{dS_{(t)}}{dt} = -\left(\beta + \beta_i\theta\right) kS_{(t)} \left(I_{w(t)} + I_{c(t)}\right) + k\gamma I_{c(t)} \\ \frac{dS_{(t)}}{dt} = k\beta S_{(t)} I_{w(t)} \\ \frac{dS_{(t)}}{dt} = \left(\beta + \beta_i\theta\right) kS_{(t)} I_{c(t)} - k\gamma I_{c(t)} \end{cases} \tag{9}$$

Suppose there is only one communicator at the beginning of the propagation of online public opinion. When t=0, the initial conditions for the spread of public opinion are as follows.

$$S_{(0)} + I_{(0)} = N,\ S_{(0)} = \frac{N-1}{N} \approx 1\left(N \to \infty\right),\ I_{(0)} = \frac{1}{N} \approx 0 \tag{10}$$

As time t increases, the propagation state of individual nodes in the network changes as follows.

$$S_{(t)} + I_{w(t)} + I_{c(t)} = 1,\ I_{w(0)} = \beta N,\ I_{c(0)} = 0 \tag{11}$$

Note that compared to the traditional SIS model, the infected person I of the SI_wI_cS network water army propagation model has two different states, the water army I_w and the infected person I_c infected by the water army.

3 $\mathbf{SI_wI_cS}$ Model analysis of internet water army

In the process of spreading online public opinion, the spreading nodes will keep increasing, and when the online public opinion reaches the peak, the water army will have the highest spreading efficiency at this time, and the spreading nodes will gradually decrease as the hotspot heat dissipates. S, I_w and I_c are used to denote the final states of susceptible, aquatic, and infected individuals, respectively. Note that $S_{(t)}$, $I_{w(t)}$ and $I_{c(t)}$ have $S_{(t)}$+$I_{(t)}$=1 for any time t and $I_{(0)}$. At this point, $I_{(t)}$ represents the sum of the water forces and the infected. The above equation demonstrates that $S_{(t)}$ and $I_{(t)}$ are the proportion of susceptible and infected people in the population, respectively, where $S_{(t)}$+$I_{(t)}$=1 ($0 < S, I \leq 1$). Now the steady state of the SI_wI_cS internet water army model is analyzed, based on the model equation of the SI_wI_cS network water army, the equation can be obtained as follows.

$$\begin{cases} \dfrac{dS_{(t)}}{dI_{w(t)}} = \dfrac{-\left(\beta+\beta_i\theta\right)kS_{(t)}\left(I_{w(t)}+I_{c(t)}\right)+k\gamma I_{c(t)}}{\beta kS_{(t)}I_{w(t)}} \\ I_{w(t)} = \dfrac{\beta kS_{(t)}I_{w(t)}}{\beta S_{(t)}\left(I_{w(t)}+I_{c(t)}\right)+\beta\beta\theta\left(\Gamma_{(t)}+\Omega_{(t)}+\pi_{(t)}\right)S_{(t)}\left(I_{w(t)}+I_{c(t)}\right)-\gamma I_{c(t)}}\left(1-S_{(t)}\right) \\ I_{w(t)} = \dfrac{\beta S_{(t)}(1-S_{(t)})+\gamma I_{c(t)}}{\beta S_{(t)}+\beta\beta\theta\left(\Gamma_{(t)}+\Omega_{(t)}+\pi_{(t)}\right)S_{(t)}} - I_{c(t)} \\ \\ \dfrac{dS_{(t)}}{dI_{c(t)}} = \dfrac{-\left(\beta+\beta_i\theta\right)kS_{(t)}\left(I_{w(t)}+I_{c(t)}\right)+k\gamma I_{c(t)}}{\left(\beta+\beta_i\theta\right)kS_{(t)}I_{c(t)}-k\gamma I_{c(t)}} \\ I_{c(t)} = \dfrac{I_{c(t)}(\beta\beta\theta\left(\Gamma_{(t)}+\Omega_{(t)}+\pi_{(t)}\right)S_{(t)}-\gamma)}{\beta S_{(t)}I_{c(t)}+\beta S_{(t)}I_{w(t)}+\beta\beta\theta\left(\Gamma_{(t)}+\Omega_{(t)}+\pi_{(t)}\right)S_{(t)}I_{c(t)}}\left(1-S_{(t)}\right) \\ I_{c(t)} = \dfrac{\left(\beta\beta\theta\left(\Gamma_{(t)}+\Omega_{(t)}+\pi_{(t)}\right)S_{(t)}-\gamma\right)\left(1-S_{(t)}\right)-\beta S_{(t)}I_{w(t)}}{\beta S_{(t)}+\beta\beta\theta\left(\Gamma_{(t)}+\Omega_{(t)}+\pi_{(t)}\right)S_{(t)}} \end{cases} \begin{matrix} (12) \\ \\ \\ (13) \end{matrix}$$

Note that for any given time t and , there is. For the purpose of the following discussion, use so that the following results are obtained.

$$S_{(t)} = e^{-\varepsilon(1-S_{(t)})} \tag{14}$$

$$I_{(t)} = 1-S_{(t)} = 1-e^{-\varepsilon\left(1-S_{(t)}\right)} \tag{15}$$

$$\varepsilon = 1+\frac{\left(\beta+\beta\theta\left(\Gamma_{(t)}+\Omega_{(t)}+\pi_{(t)}\right)\right)\beta+\gamma}{\left(\beta+\beta\theta\left(\Gamma_{(t)}+\Omega_{(t)}+\pi_{(t)}\right)\right)\gamma} \tag{16}$$

Eq(15) is the same as the transcendental equation for SIR rumor propagation (Moreno et al (2004)). According to this theorem obtained by Zan et al.(2014) as follows. For $\varepsilon > 1$, equation X $= e^{-\varepsilon(1-X)}$ has two solutions; X_1=1 and a nontrivial solution X_2, where $0 < X_2 < 1 - \frac{\ln\varepsilon}{\varepsilon}$. $I_{(t)}$=1-$S_{(t)}$ always allows a nontrivial solution S_1=1 also since $\varepsilon > 1$ all nontrivial values of β and β_i can be realized. At this point, $I_{(t)}$=1-$S_{(t)}$ still has another solution S_2 where $0 < S_2 < 1 - \frac{\ln\varepsilon}{\varepsilon}$. This demonstrates that there is no threshold for public opinion in the SI_wI_cS water army model. However, another indicator used to measure the maximum influence of public opinion is the peak of the I_{max}. For the peak of I_{max}, the sum of I_w and I_c in this paper is taken as the peak of the online water army for the spread of online public opinion. Therefore, from the above equation, it can be obtained that the key factor to measure the influence of the water army on the online public opinion is the peak of the spread of the water army, and the study by QIU (2018) shows that the enhancement and weakening of the heat of the public opinion can be visually reflected by the movement of the people of the spreaders (water army). Let the public opinion fervor enhance when $\frac{\partial I}{\partial t} > 0$ and weaken when $\frac{\partial I}{\partial t} < 0$. Let the left side of the equation after adding the dynamic factor be equal to zero. The following relationship equation can be obtained.

$$\frac{dI_{(t)}}{dt} = k\beta S_{(t)}I_{w(t)}+\left(\beta+\beta_i\theta\right)kS_{(t)}I_{(t)}-k\gamma I_{c(t)} = 0 \tag{17}$$

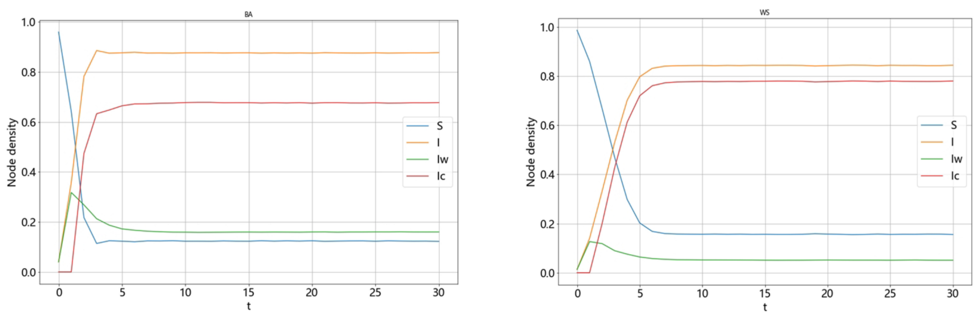
$$\frac{I_{w(t)}}{I_{c(t)}} = \frac{\gamma-(\beta+\beta\theta(e^{\frac{\mu}{t+1}}+e^{\frac{(t-X)^2}{\delta}}+\log_{10}(\frac{I_{(t)}+p}{e^2})))}{\beta S_{(t)}} \tag{18}$$

From the above equation, it can be seen that the value of $\frac{I_{w(t)}}{I_{c(t)}}$ is in an increasing state as time increases, thus this paper sets an $I^*_{(t)}$ such that $I^*_{(t)} = \frac{I_{w(t)}}{I_{c(t)}}$. From this, it can be obtained that $I^*_{(t)}$

is used as a turning point for the spread of public opinion in social networks, and the topic of public opinion will continue when $I_{(t)} > I^*_{(t)}$ and will weaken when $I_{(t)} < I^*_{(t)}$.

4 Experimental simulation and analysis

Assume that the network size at the beginning is N=5000, and assume that the initial state of the network has only susceptible nodes and a predetermined number of nodes of the water army, and let the number of water army at the beginning be 500. Assume that the node of the sailor is 100% infected for the neighboring nodes. The effective propagation rate at the beginning β=0.5, and the susceptible node is infected by the water army to become infected. As time increases, the propagation efficiency of the infected node changes to β_i after encountering experiencing 3 effects, with an infection rate of θ=0.3 for neighboring nodes. It will eventually change back to the susceptible node state with a probability of γ=0.3. The relevant experimental parameters are set to X=1, δ=20, p=1, μ=1, the average degree k=5, and the number of iterations is 30 according to the experimental needs. As shown in Fig.3, Fig.3(a) and Fig.3(b) represent the internet water army model of SI_wI_cS network under BA scale-free network and WS small-world network, respectively.



(a) (b)

Fig. 3 Scale-free and small-world networks with a water army model

In Fig.3, S represents the susceptible person, I_w represents the node of the water army, I_c represents the node of the infected person infected by the water army, and I represents the sum of the number of nodes of I_w and I_c, which is the total number of propagation.

After analyzing the curve changes in Fig.3(a) and Fig.3(b), the following conclusions can be obtained. In the initial population the water army is present at the beginning, and as the Internet opinion spreads, the number of susceptible people is decreasing, while the proportion of water army and infected people is increasing. At the beginning of the first round of transmission, the water army infects susceptible people and turns them into infected people, at which time the number of nodes in the water army rises rapidly and reaches its peak value, the susceptible people are separated from the water army and the water army reverts to the original number of nodes. The probability of infection and the number of network nodes do not change, while the probability of propagation of the infected person under different state effects will change dynamically, and will change to the susceptible state with a certain probability as time increases.

In the plots of BA scale-free and WS small-world networks, the peak of the spread of the water army opinion can be seen. Time t is between 0 and 5, when the spread of the water army reaches its highest point and the infected person follows it.

As the population of susceptible people continues to decrease, the water army continues to spread public opinion with a certain probability. The infection probability of the infected person is constantly changing, and the infected person keeps infecting the susceptible person on the one hand, and becomes susceptible with a certain probability on the other hand. In order to better describe the influence of online water army and dynamic effect on the probability of spreading public opinion, the probability of spreading public opinion is its initial probability of spreading.

4.1 Different initial propagation rates

The initial propagation probability β was changed under the BA scale-free network and WS small-world network starting from 30 iterations when making it 0.1, 0.2, 0.4, 0.6, and 0.8, respectively, and the other data were kept constant. The following Fig.4 is obtained. Fig.4(a) and Fig.4(b) represent

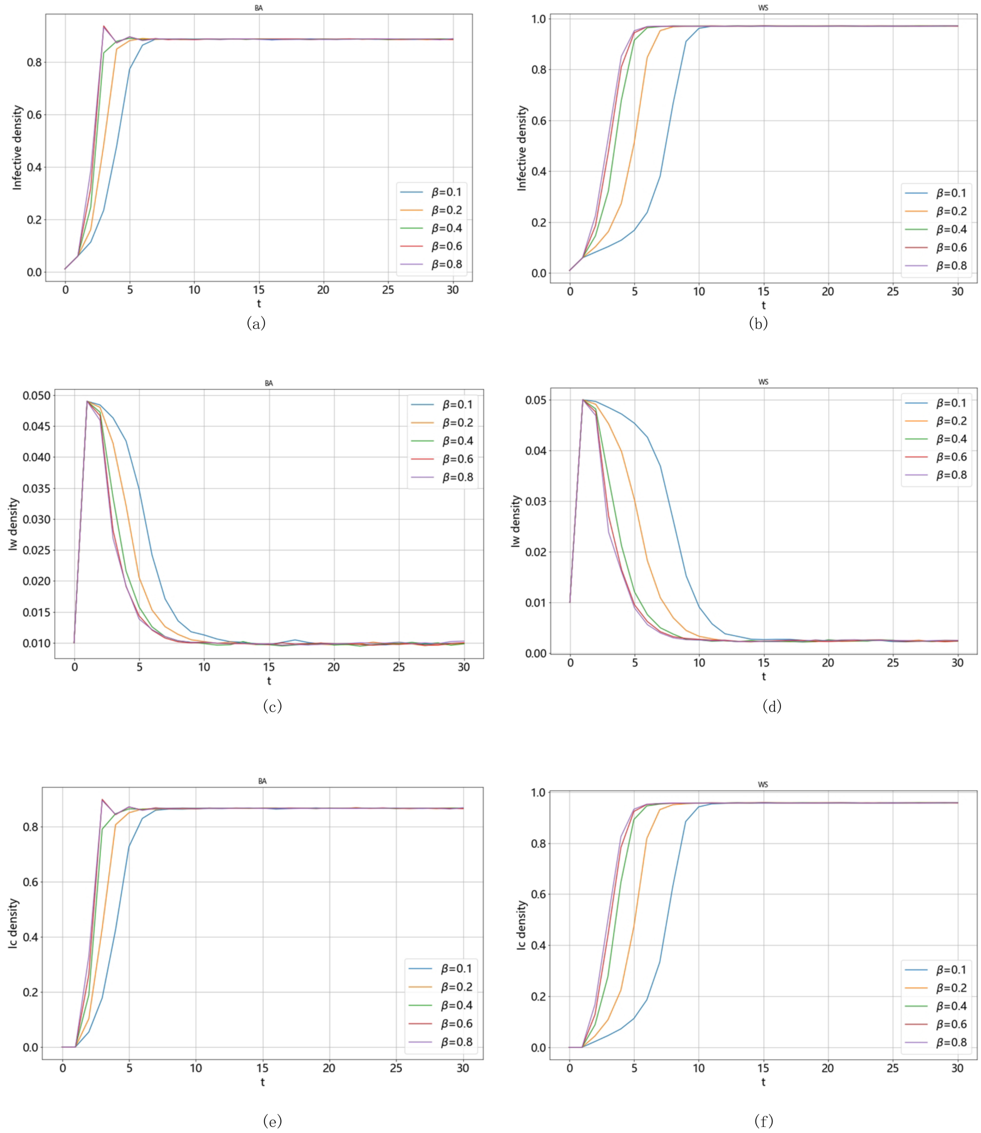


Fig. 4 Effect of different initial propagation rates on propagation

the propagation trends of the sum of water army (the number of water army plus the number of infected people) for online public opinion under different initial propagation probabilities in the BA scale-free network and WS small-world network, respectively. From Fig.4(a) and Fig.4(b), it can be seen that the propagation rate increases when the initial propagation probability becomes larger. The greater the dissemination rate, the faster the rate of dissemination and the shorter the time to reach the peak of public opinion.

Fig.4(c) and Fig.4(d) represent the effects of different initial propagation probabilities on the propagation rate of the water army under BA scale-free network and WS small-world network, respectively. From Fig.4(c) and Fig.4(d), we can see that different initial propagation probabilities have different effects on the propagation rate of the sailors. The larger the initial propagation probability, the faster the rate of propagation and the shorter the time to reach the peak of public opinion. The increase in the number of water army in the figure is due to the addition of infected people from the first round of infection, and as time increases the water army reverts to the original number of nodes and spreads with a certain probability of infection. From Fig.4(c) BA scale-free network it can be seen that the rate of propagation is more rapid than in Fig.4(d) WS small world network.

Fig.4(e) and Fig.4(f) represent the effect of different initial propagation rates on the propagation rate of infected individuals under BA scale-free network and WS small-world network, respectively. From Fig.4(e) BA scale-free network, it can be seen that the larger the initial propagation rate is the faster the propagation rate. The same is true in Fig.4(f) WS small-world network, where the larger the initial propagation probability, the faster the rate of propagation. From Fig.4(e) and Fig.4(f), it can be seen that under the same initial propagation rate, the rate of propagation in Fig.4(e) BA scale-free network graph is faster than that in Fig.4(f) WS small-world network graph.

The following conclusion can be drawn from Fig.4, when other factors do not change, when we change the initial propagation probability β and randomly select a propagation node for 30 iterations of the simulation study will find that as the value of the initial propagation probability β is increasing, the propagation rate is also accelerating. In this comparison experiment, the initial propagation probability is chosen in an order from small to large, and the maximum value of dynamic propagation is set to 0.8, which is to form a sharp contrast state. From Fig.4, it can be seen that β=0.8, the rate of propagation is the fastest, and the range of propagation is also the largest. The larger the value of the initial propagation rate, the faster the propagation rate. Under the influence of three different dynamic effects, the greater the initial propagation probability of network water army, the faster the spread of network public opinion, the deeper the degree of diffusion, and the wider the scope of diffusion.

4.2 The influence of communicator importance on communication

In a complex network, the importance of an individual is measured by the degree of the node it represents. In general, the greater the degree of a node, the higher its iortance. The experimental simulation of BA scale-free network and WS small-world network was performed with 5000 nodes with the minimum degree of 5 and the maximum degree of 50. For simplicity of operation, k nodes of 5, 10, 20, 30, 40, and 50 were selected as the initial propagators, and the rest of the values were kept constant. 30 iterations were performed, and six consecutive experiments for each case were averaged, and then the propagation density was obtained with the The curve of propagator variation. The curve is shown in Fig.5.

Fig.5(a) and Fig.5(b) represent the effects of different degree sizes of nodes on the propagation trend of online public opinion under the total number of online water army nodes (the sum of the number of water army and infected people) in the BA scale-free network and WS small-world network, respectively. Regardless of whether it is under the BA scale-free network in Fig.5(a) or the WS small-world network in Fig.5(b), when the degree of the nodes is larger, the faster for the propagation rate of public opinion and the shorter time to reach the peak.

Fig.5(c) and Fig.5(d) represent the effect of different node degree sizes on the propagation rate of the sinker nodes under the BA scale-free network and WS small-world network. The larger the degree of a node in the same time period, the faster the rate of propagation and the stronger the degree of infection for surrounding neighboring nodes. Over time, the nodes infected by the water army are turned into the state of the infected, and the infected will change to the state of the susceptible with a certain probability. The state of the water army will change back to its initial state, and its number and propagation probability will remain unchanged, continuing to spread the infection to the susceptible nodes.

Fig.5(e) and Fig.5(f) represent the effects of different node degree sizes on the spread of online opinion by infected people under BA scale-free network and WS small-world network, respectively. From Fig.5(e) and Fig.5(f), we can see that different degree sizes of the same node are different for the rate of propagation. The larger the degree of the node, the faster the propagation rate, the larger the infection range for the surrounding neighboring nodes, and the shorter the time to reach the peak of propagation, and the more rapidly for the propagation of public opinion.

The conclusion can be drawn from Fig.5. When the importance of the communicator is greater, it spreads faster and faster, and the scope of the spread is larger, and the time to reach the peak is shorter, and it continues to spread with a certain probability after reaching the peak. The importance of the initial communicator's meal can affect the spread of public opinion and the speed of spreading, and the greater the importance of the communicator, the faster the spread of public opinion and the wider the spread of public opinion.

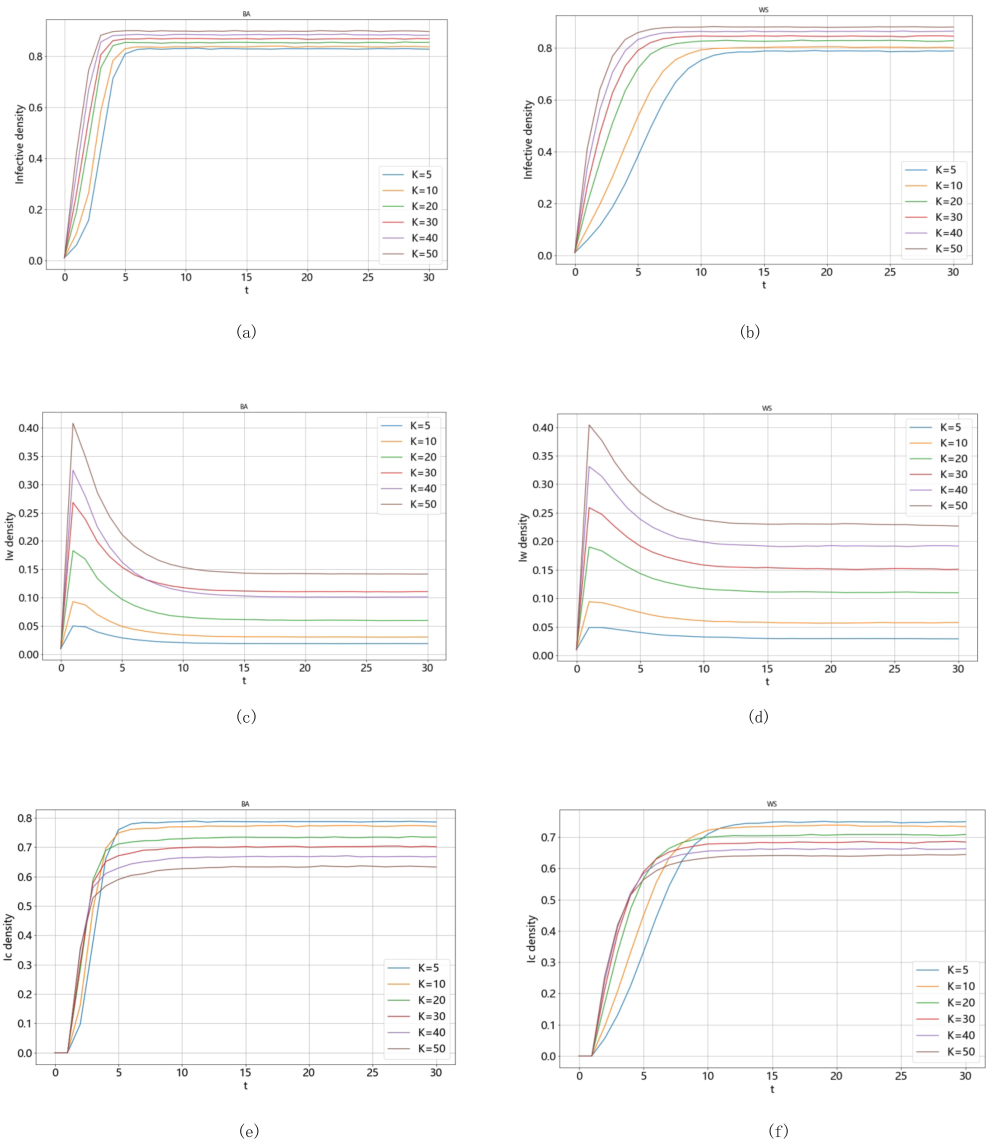


Fig. 5 Influence of communicator importance on the spread of online public opinion

4.3 Different water army proportion spread comparison

To explore the importance of online water army for the spread of online public opinion, summarize the number of points N=1000, and set different proportions of water army I1(10%), I2(20%), I3(30%), I4(40%), and I5(50%) in the population in the state of initial spread with other conditions unchanged, and observe the impact of water army on the spread of public opinion. As shown in Fig.6.

Fig.6(a) and Fig.6(b) represent the impact of different proportions of initial water soldiers on the propagation of online public opinion in BA scale-free network and WS small-world network, respectively. It can be clearly seen in Fig.6(a) and Fig.6(b) that when the initial proportion of water army is larger, the stronger the propagation for online public opinion is. In the same time period, the time for public opinion to reach the peak is almost the same, but the higher the initial proportion of the water army, the higher the height of the peak and the wider the spread.

4.4 Influence of Water Army on Public Opinion Dissemination

To investigate the influence of online water army on the spread of online public opinion, the experimental simulation is set up with two different comparison tests, and the probability of spread and the total number of nodes of the crowd at the beginning are kept the same. The number of summation points is set to N=5000, and other data are kept constant. The first group is called the water army group I1, 500 nodes were pre-selected as the original water army nodes at the beginning

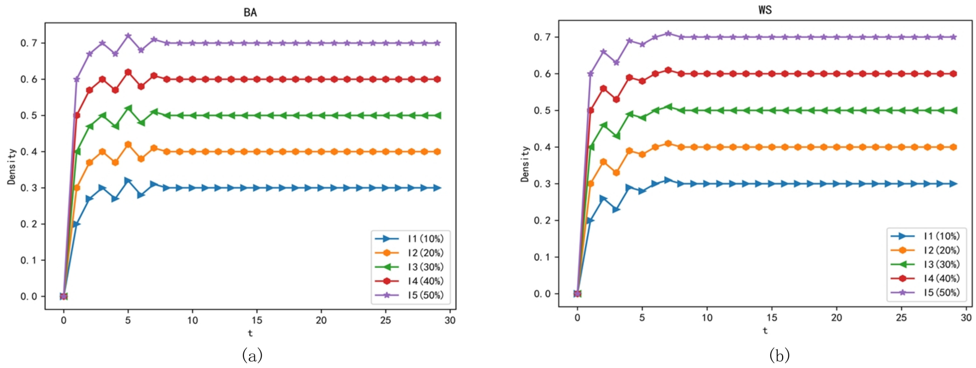


Fig. 6 Effect of the proportion of initial water forces on the spread of public opinion

of the experiment to simulate the spread of the water army in the network. The second group, called the general group I2, does not select any nodes as the sailors at the beginning of the experiment, but only 500 nodes as the infected, and simulates the propagation of public opinion in the network with the propagation form of the infected, and the experiment is iterated 30 times. This is shown in Fig.7.

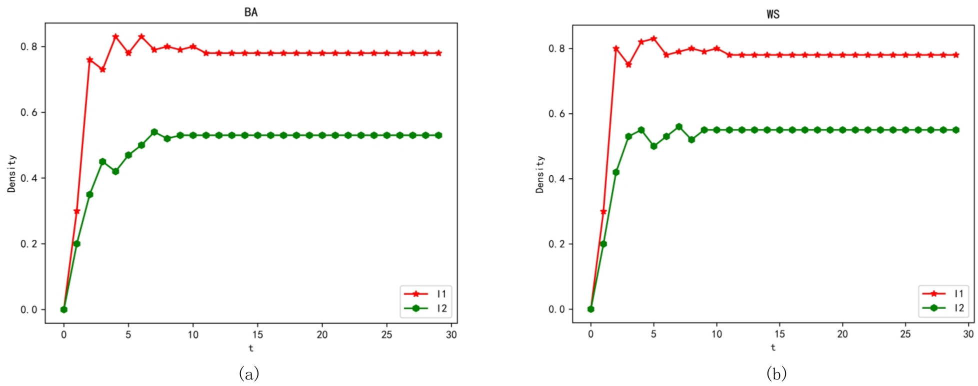


Fig. 7 Comparison of the propagation rate of public opinion by water army and ordinary communicators

Fig.7(a) and Fig.7(b) represent the comparison of online opinion spreading with and without online water army in BA scale-free network and WS small world network, respectively. From Fig.7(a) and Fig.7(b), it is obvious that at the beginning of public opinion spreading, it is clear that the spreading of public opinion becomes faster and more widespread after the inclusion of online water army. In the network where no water army is added, it can be seen that the speed of spreading is not as fast as in the network where water army is added, and the scope of spreading is not as large as in the network where water army is added. Thus, it can be seen that the water army has a certain influence on the spread of online public opinion.

5 Real Data Acquisition and Model Comparison

5.1 Real data on Weibo Internet water armies

In view of the characteristics of information dissemination of internet water army, how to identify internet water army in social networks is a key point of research in this paper. Using crawler technology, we crawl the corresponding Weibo data, and use the number of retweets and comments of that Weibo at that time as the basis for screening internet water army, and draw on the information propagation model based on interaction behavior proposed by Chen et al.(2015) The real data of Sina Weibo were used to analyze the propagation model and validate the detection method, and the results showed that the detection method could effectively detect the information of water army in Weibo. The study by Bindu et al.(2018) shows that most of the internet water army retweets use two “guidance tools” provided by Sina Weibo, namely the topic symbol “#” and external link URLs, in order to categorize user Weibo content and strengthen the purpose of third-party connections. In general, normal users do not use these two lead generation tools on a large scale because of their own social diversity. However, the water army often uses these two guidance tools on a large scale in order to achieve the purpose of shaping public opinion such as topic gathering and heat generation.

5.1.1 The number of characteristics of weibo water army

The data between February 20th and February 25th were counted with the corresponding topic symbol “#” and external link URLs for the number of retweets on the tweets, and a total of 10,645 data were obtained. This is shown in Fig.8.

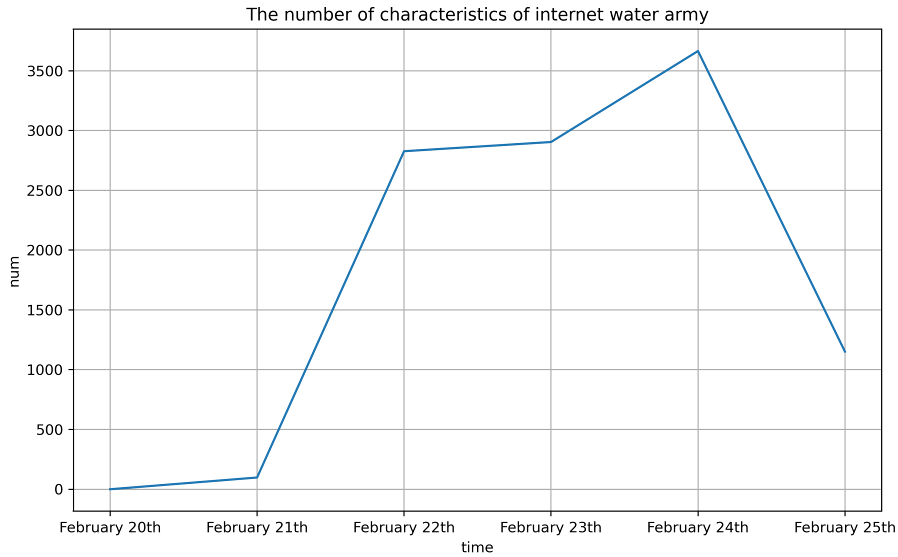


Fig. 8 Internet water army characteristic number

5.1.2 The timeliness of Weibo hotspots

The UIDs of one original Weibo post (17498a725d6459c 72d2fc4ble98e5ee) of the “Cargo Lala incident” were tracked and filtered. After the event, between February 20th and February 25th, the number of retweets by different users was counted, and the tweets with external link URLs were filtered together, and the retweets were counted accordingly, and a total of 4125 data were obtained for visualization and analysis, and Fig.9 was obtained.

From Fig.9, we can conclude that, for an original Weibo, tracking the UID of its root Weibo, we found that both using the UID and URL of a root Weibo for retweeting makes the internet water army have a peak of public opinion spreading, and the Weibo reached a peak on February 22th. With different public opinion information, the amount of retweets for this Weibo gradually decreased. At the same time, it can be seen that the retweets of a Weibo are time-sensitive, generally maintaining a heat level of about 3 days.

5.1.3 Schematic diagram of the spread of internet water armies

In order to better observe the influence of online water army on the spread of online public opinion, we analyzed the number of retweets (containing URL or adding the “#” retweet flag) that water army started to make for different root Weibo UIDs on the same day for the “Cargo Lala incident”. In Fig.10(a), the white nodes are the nodes of the water army, and the red ones are the nodes of the susceptible. The nodes in white in Fig.10(b) include the water army nodes and the susceptible nodes infected by the water army. As shown in Fig.10.

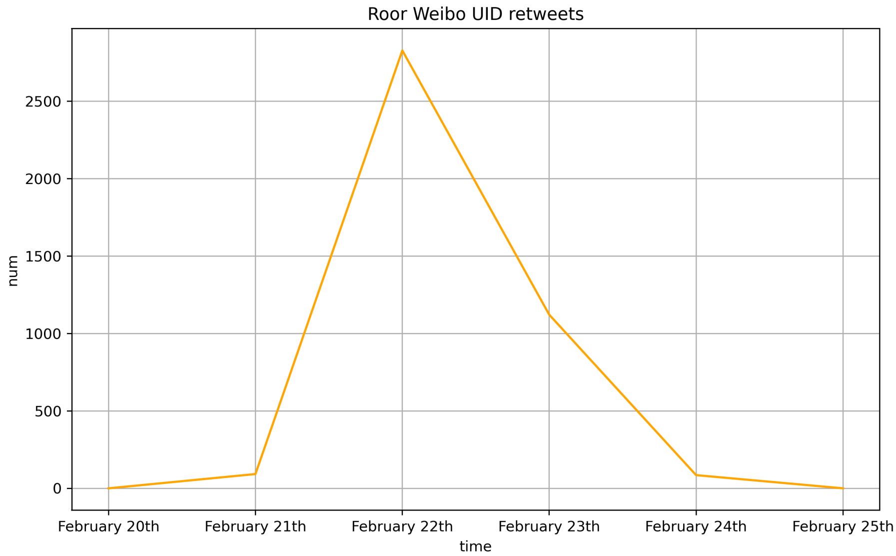
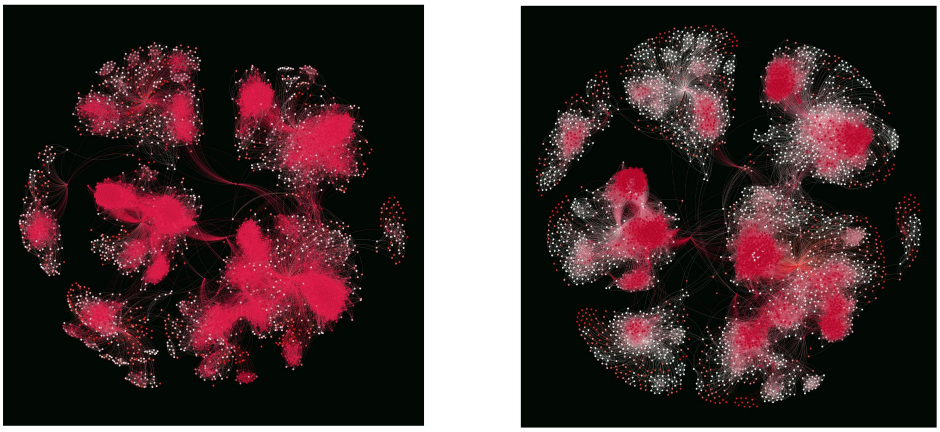


Fig. 9 Number of retweets from the root Weibo UID



(a) (b)

Fig. 10 Internet water army node propagation state diagram

From Fig.10(a), it can be seen that in the initial state the crowd selected a part is as a water node to infect the susceptible nodes for spreading the public opinion. Fig.10(b) shows the state of the nodes after 30 rounds of infection. It can be seen from the Fig.10(a) that the susceptible node has been infected by the node of the water army and the infected node under the joint propagation of the node state. From the Fig.10(b), it can be seen that after many rounds of propagation, the nodes of water army and infected person occupy most of the nodes. Fig.10(b) shows that as time increases, the susceptible nodes have been infected more and more by the nodes of the water army and the infected nodes, and the public opinion has become diffuse, and the direction of public opinion in the whole population has been gradually dominated by the water army and the infected.

5.2 Comparison of experimental data with real data

5.2.1 Real Web water armies Spreading Online Opinion

Currently, the World Wide Web has become a carrier of massive amounts of data. Crawling, analysis and statistics of large amounts of data have become important tools for Internet research, and web crawlers facilitate this task. The web crawler was used to crawl the data related to the

"Cargo Lala jumped out of the car" incident that occurred in 2021 on Sina Weibo. The main objects crawled are the number of retweets and comments about the event, as well as the topic symbol "#" and external link URLs. As shown in Fig.11.

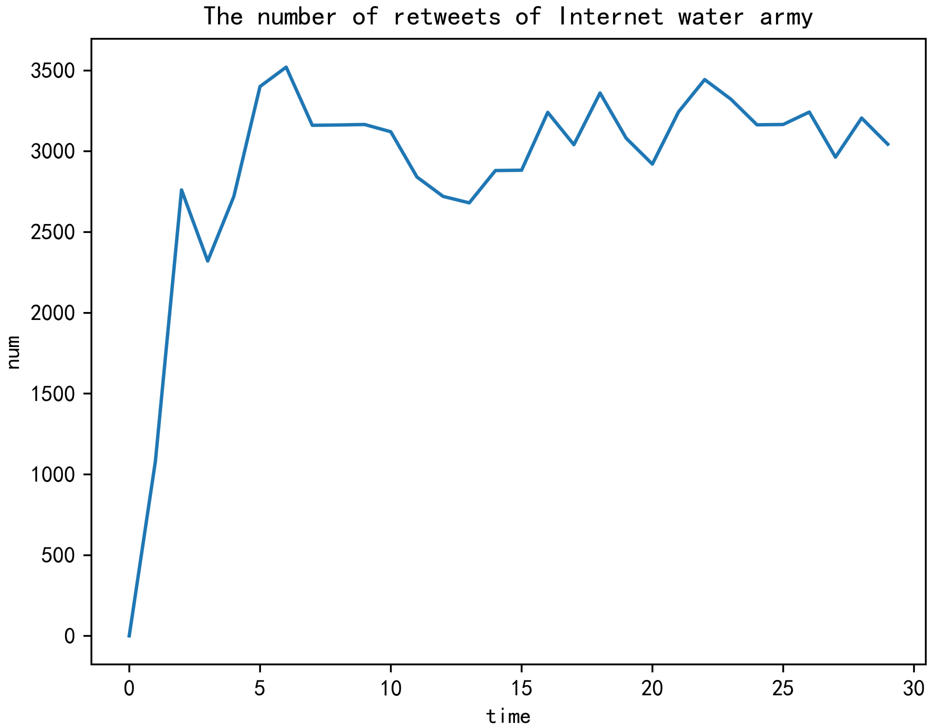


Fig. 11 Retweets of Weibo by internet water army

Fig.11 shows the retweets of the Weibo "Cargo Lala incident" from February 20th and February 25th, filtering the number of retweets and comments less than 100 or more, as well as the number of retweets with the topic symbol "#" and external links. We also filtered the number of retweets with topic symbols "#" and external links to Weibos or Weibos containing "#" for retweets, and finally obtained 857 Weibo data sets about water army, with 410,995 retweets. In the beginning, the number of retweets gradually increased with time, so the retweets on the February 20th day started from zero, after which the number of retweets screened each day was required to be more than 1000 retweets, and the retweets also had external link URLs to Weibos or Weibos containing "#" that matched the identity of the sailors. Special characters, through the retweet volume of these Weibos, we can see about the influence of online water army on the spread of online public opinion. At this point, the retweets of the water army represent the retweets of the water army and the infected person together for the Weibo.

5.2.2 Comparison of experimental data with real data

Fig.12 shows the comparison between the real data of Weibos (data from Sina Weibo Hotspot) and the simulated data. The curves in the figure represent the propagation trends of online water army on online public opinion in BA scale-free network and WS small-world network for real data and experimental simulated data of Weibos, respectively. In Figure 12(a) and Figure 12(b), we can see that the experimental simulation data in the SI_wI_cS internet water army model is very close to the trend of the spread of internet water army in the real data of the "Cargo Lala incident" in Weibo. From Figure 12(a) and Figure 12(b), we can see that at the beginning, as the internet water army spreads to the public opinion, the density of infection is increasing, but as the impulse effect, the heat effect and the herd effect diminish, the spread rate of the internet water army gradually tends to level off. Compared with the real data, there will be some ups and downs, and finally the overall tendency to level off, which is consistent with the SI_wI_cS internet water army model propagation

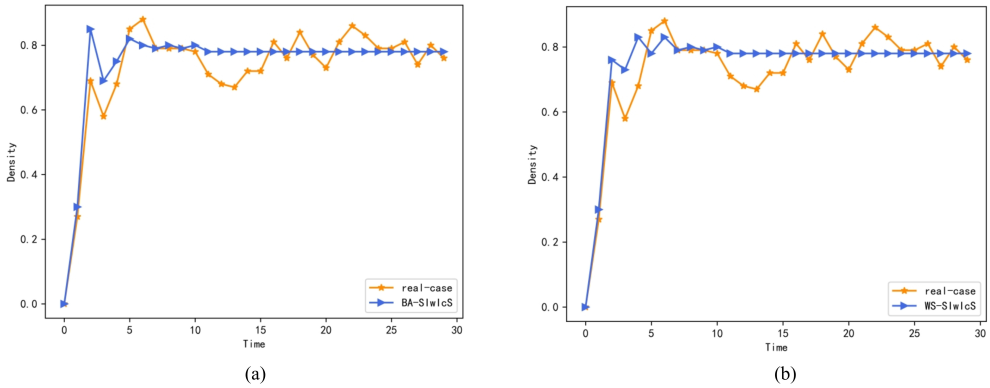


Fig. 12 Comparison of simulated experimental data and real data of Weiboging

direction. The results obtained through experiments and real data comparison tests prove that the SI_wI_cS internet water army model proposed in this paper is with real reliability.

6 Conclusion

In this paper, by establishing the SI_wI_cS internet water army dissemination model, we got the trend process about the dissemination trend of Internet water army on Weibos about Internet public opinion, and studied to get the process of how the Internet water army and the ordinary communicators spread Internet public opinion together in the network. By filtering the number of retweets and comments on the corresponding Weibo, as well as the Weibo data that match the topic symbol "#" and the URL of the external link, the data set that match the identity of internet water army is obtained. Using the real data of internet water armies in Sina Weibo to analyze the propagation model and verify the feasibility of the SI_wI_cS Internet water armies model, the results show that the model in this paper can efficiently predict the trend of Internet water armies for the propagation of Internet public opinion. Although there are differences in the types and functions of Internet water armies, the commonality of communication behaviors makes the SI_wI_cS Internet water army communication model more versatile for the communication of online public opinion on Weibo, and it can also be applied to the prediction of Internet water army's dynamic communication of online public opinion in multiple scenarios.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.62266030, No.61863025), the Program for International S & T Cooperation Projects of Gansu province (No.144WCGA166), Program for Longyuan Young Innovation Talents and the Doctoral Foundation of LUT. We would like to thank the second "Weibo Hotspot" data mining competition for supporting this article on the dataset.

Author contributions statement

Fuzhong Nian: conceptualization, methodology. Chongpei Wang: data curation, writing original draft preparation, software. Duan Zhang,investigation, supervision. Zhongkai Dang: analysed the results. All authors reviewed the manuscript.

Competing interests

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.

References

Bathla G, Singh P, Singh RK, et al (2022) Intelligent fake reviews detection based on aspect extraction and analysis using deep learning. Neural Computing and Applications 34(22):20213–20229

Bindu P, Mishra R, Thilagam PS (2018) Discovering spammer communities in twitter. Journal of Intelligent Information Systems 51:503–527

Chen C, Wu K, Srinivasan V, et al (2013) Battling the internet water army: Detection of hidden paid posters. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp 116–120

Chen G, Cai W, Huang J, et al (2016) Uncovering and characterizing internet water army in online forums. In: 2016 IEEE First International Conference on Data Science in Cyberspace (DSC), IEEE, pp 169–178

Chen K, Chen L, Zhu P, et al (2015) An online social network water army detection method based on interactive behavior [j]. Journal of communications 36(7):120–128

Chen M, Du W (2023) The predicting public sentiment evolution on public emergencies under deep learning and internet of things. The Journal of Supercomputing 79(6):6452–6470

Ebbinghaus H (2013) Memory: A contribution to experimental psychology. Annals of neurosciences 20(4):155

Geng L, Zheng H, Qiao G, et al (2023) Online public opinion dissemination model and simulation under media intervention from different perspectives. Chaos, Solitons & Fractals 166:112959

Guo B, Jiang Zb, et al (2023) What is the internet water army? a practical feature-based detection of large-scale fake reviews. Mobile Information Systems 2023

He S, Hollenbeck B, Overgoor G, et al (2022) Detecting fake-review buyers using network structure: Direct evidence from amazon. Proceedings of the National Academy of Sciences 119(47):e2211932119

He Y, Yang P, Cheng P (2023) Semi-supervised internet water army detection based on graph embedding. Multimedia Tools and Applications 82(7):9891–9912

Jabeur SB, Ballouk H, Arfi WB, et al (2023) Artificial intelligence applications in fake review detection: Bibliometric analysis and future avenues for research. Journal of Business Research 158:113631

Jain A, Dhar J, Gupta VK (2023) Optimal control of rumor spreading model on homogeneous social network with consideration of influence delay of thinkers. Differential Equations and Dynamical Systems 31(1):113–134

Jiaze X (2021) Analysis on the influence of water army on public opinion on the internet. In: 2021 3rd International Conference on Literature, Art and Human Development (ICLAHD 2021), Atlantis Press, pp 718–722

Lee M, Song YH, Li L, et al (2022) Detecting fake reviews with supervised machine learning algorithms. The Service Industries Journal 42(13-14):1101–1121

Li D, Li Q, Hu Y, et al (2014) An approach to detect the internet water army via dirichlet process mixture model based gsp algorithm. In: Applications and Techniques in Information Security: 5th International Conference, ATIS 2014, Melbourne, VIC, Australia, November 26-28, 2014. Proceedings 5, Springer, pp 82–95

Liu W, Xia Z (2021) Detect network water army based on large-scale e-commerce reviews. In: 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), IEEE, pp 273–278

Moreno Y, Nekovee M, Pacheco AF (2004) Dynamics of rumor spreading in complex networks. Physical review E 69(6):066130

Peng J, Wang Y, Meng Y (2023) Detecting e-commerce water army through graph modeling on user multiple collusive relationships: A case study of china's hotel industry. Journal of Theoretical and Applied Electronic Commerce Research 18(1):105–129

SHEN Q, GAO Y, NIU L (2017) Based on the tensor flow framework and the improved logistic regression algorithm, the detection model of weibo water army was constructed. In: International Conference on Computer Engineering,Information Science and Internet Technology(CII 2017)

Shin Y, Gupta M, Myers S (2011) Prevalence and mitigation of forum spamming. In: 2011 Proceedings IEEE INFOCOM, IEEE, pp 2309–2317

Vidanagama D, Silva A, Karunananda A (2022) Ontology based sentiment analysis for fake review detection. Expert Systems with Applications 206:117869

Wang J, Li Y (2023) Research on the propagation and governance of public opinion information under the joint action of internal and external factors. Aslib Journal of Information Management 75(2):193–214

Wang K, Xiao Y, Xiao Z (2014) Detection of internet water army in social network. In: 2014 International Conference on Computer, Communications and Information Technology (CCIT 2014), Atlantis Press, pp 189–192

Wu Y, Wang D, Ma F (2023) A study on the competitive dissemination of disinformation and knowledge on social media. Aslib Journal of Information Management

Xu C, Zhang J, Chang K, et al (2013) Uncovering collusive spammers in chinese review websites. In: Proceedings of the 22nd ACM international conference on Information & Knowledge Management, pp 979–988

Yan S, Su Q, Wu L, et al (2023) A damping grey multivariable model and its application in online public opinion prediction. Engineering Applications of Artificial Intelligence 118:105661

Zeng K, Wang X, Zhang Q, et al (2014) Behavior modeling of internet water army in online forums. IFAC Proceedings Volumes 47(3):9858–9863

Zhang H, Li T, Wang Y (2019) A social network water army detection model based on artificial immunity. In: Journal of Physics: Conference Series, IOP Publishing, p 052097

Zhang Z, Mei X, Jiang H, et al (2023) Dynamical analysis of hyper-sir rumor spreading model. Applied Mathematics and Computation 446:127887