future internet

Article

A Game Theory-Based Model for the Dissemination of Privacy Information in Online Social Networks

Jingsha He, Yue Li and Nafei Zhu *

Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
* Correspondence: znf@bjut.edu.cn; Tel.: +86-10-67396272

Abstract: Online social networks (OSNs) have experienced rapid growth in recent years, and an increasing number of people now use OSNs, such as Facebook and Twitter, to share and spread information on a daily basis. As a special type of information, user personal information is also widely disseminated in such networks, posing threats to user privacy. The study on privacy information dissemination is thus useful for the development of mechanisms and tools for the effective protection of privacy information in OSNs. In this paper, we propose to apply the game theory to establish a sender–receiver game model and the Nash equilibrium to describe the behavioral strategies of users in disseminating privacy information. Factors that affect the dissemination of privacy information are also analyzed with two important aspects: intimacy and popularity of the privacy-concerning subject. Simulation experiments were conducted based on real data sets from scale-free networks and real social networks to compare and analyze the effectiveness of the model. Results show that the proposed game theory is applicable to the privacy information dissemination model, which implements intimacy and popularity in the modeling of the dissemination of privacy information in OSNs. Both the impact of the macro-level OSNs and the micro-relationships between users are evaluated on the dissemination of privacy information, which provides a new perspective for exploring the dissemination of privacy information and facilitates the development of effective mechanisms for privacy protection in OSNs.

Keywords: privacy information dissemination; game theory; Nash equilibrium

check for updates

Citation: He, J.; Li, Y.; Zhu, N. A Game Theory-Based Model for the Dissemination of Privacy Information in Online Social Networks. *Future Internet* **2023**, *15*, 92. https://doi.org/10.3390/fi15030092

Academic Editor: Carlo Blundo

Received: 9 January 2023
Revised: 23 February 2023
Accepted: 23 February 2023
Published: 27 February 2023

1. Introduction

With the advancement of information technology, the interconnection structure of the Internet has experienced rapid growth. The number of users in online social networks (OSNs) has maintained steady growth, and the level of activity of users has increased as well [1]. According to the “Global Digital Report 2021” released by We Are Social, there are currently about 4.66 billion users worldwide, among which 4.2 billion are social media users [2]. As the largest OSN platform in the world, Facebook has continued to expand its user base, contributing to the accelerated spread of information on the Internet. Meanwhile, social media users are facing increasingly more threats to information security and privacy violations [3]. Therefore, the prevention of malicious and widespread privacy information dissemination has become an urgent issue.

Although increasingly more attention has been paid to digital privacy protection on the Internet, research on privacy protection is in its early stage. The majority of the research on the dissemination of information has mostly focused on the analysis of the trend of the dissemination of public opinions. While the research on privacy information dissemination in OSNs is still in its early stage, most of the research is focused on the macro level of the network, while little is on the micro level of user psychology. In addition, little research has applied the game theory to study the dissemination of privacy information, to the best of our knowledge, nor has it used any game theory model to describe the micro-level factors regarding the dissemination of privacy information.

The aim of this paper is to study how privacy information disseminates in OSNs by taking into account the impact of micro-relationships between users, transforming different types of relationships into specific parameter indicators, and developing a game theory-based model to facilitate the study of the dissemination of privacy information. Our work also includes carrying out some experiments to evaluate the game theory-based model and validate the model using some real data from the Facebook network. Experimental results show that the game theory-based model that we proposed in this paper can be effectively used to study the dissemination of privacy information in OSNs. Such studies can help better understand the ways or patterns of privacy information dissemination in OSNs, which is important for the development of effective mechanisms or policies for the protection of privacy during the sharing of information in OSNs.

The remainder of this paper is organized as follows. Section 2 reviews some related work. Section 3 analyzes the factors that influence privacy information dissemination and introduces the implementation process, game setup process, and specific structure of the privacy information dissemination model. Section 4 explores the influence of privacy information dissemination rules through comparative experiments and analysis to verify the effectiveness of the proposed model. Section 5 concludes this paper with some suggestions for future research.

2. Related Work

Complex networks have been used in many different contexts: physics, management, biology, computing science, and sociology [4]. In 1998, Watts and Strogatz proposed a small-world network with high clustering coefficients and a small average path length [5]. In 1999, Barabasi found that the degree distribution of data in the World Wide Web obeys the power-law distribution through statistical analysis and proposed the concept of scale-free networks [6]. Such scale-free networks can be constructed using the mechanism of "optimal connection" in the network model. These network models are the foundation of complex network research. At present, many researchers mostly study the topological structure of social networks based on scale-free networks.

The game theory model is a formal model that describes the strategic decisions taken by individuals that interact with each other [7]. In 1944, the book "Game Theory and Economic Behavior" [8] written by the famed mathematician Neumann and economist Morgenstem started the discipline of game theory. In 1950, Nash proposed the concept of non-cooperative game strategy equilibrium, which is called the Nash equilibrium [9]. The basic elements in the game theory include the players who are the main bodies participating in the game activity and the strategy that each player can choose. The set is the game strategy and the player's income when adopting different game strategies is the game profit. As game theory focuses on the interactive participation strategies of game participants, it can describe the decision-making and selection of nodes in the face of different situations to provide theoretical support for the construction of information dissemination models. Games are often used in combination with complex networks to solve problems in such networks. Chen et al. presented a structural discrete choice model with social influence for large-scale social networks, which is based on an incomplete information game that allows individual-specific parameters of the players [10]. Weyrich et al. used serious games to simulate and test how public information from social media is used in emergency operation centers to make decisions [11]. Gorelov studied two problems of rational information aggregation in hierarchical games [12]. Simon researched myopic equilibria based on games of incomplete information [13]. Aimed at the problem of the influence of node attitude on information dissemination, Huang et al. proposed an information propagation model based on an evolutionary game [14].

The process of information reaching individuals through interactive behavior is called information dissemination. Lai et al. presented a message propagation model over social networks by analyzing the relationships among nodes [15]. Hartmann explained viral message propagation in social media [16]. Gu et al. suggested that personal willingness is one of the most important factors that influence the construction of social community and message propagation in social networks, which is used to describe the subjective initiative of users to exchange information with the outside world [17]. Brusco et al. described the basic principles of affinity propagation and its relationships to the clustering problem [18].

There has also been some other research on privacy protection in social networks. Bi proposed an aggregation encryption method on social network privacy data based on matrix decomposition [19]. Adjei et al. examined the factors that influence user's decision decisions on disclosing personal information on social media and their antecedents [20]. Li and Zeng presented a novel network representation learning model to generate node embedding that can afford data incompleteness that comes from user privacy protection [21]. Bioglio and Pensa extended the SIR (susceptible infective removal) model to analyze the influence of user's privacy awareness on the dissemination of privacy information [22] and how neighbor nodes would affect the dissemination of privacy information in OSNs [23].

However, most of the previous research on OSNs has been focused on the spread of information and rumors in OSNs, and the research on private messages has not adequately considered user's micro-influence. Meanwhile, the game theory model incorporated in the social networks is a pure strategy game scenario of a single mode. At present, there is hardly any research on mixed games between the sender and the receiver.

Based on the influencing factors of the relationships between users, it is possible to determine the probability that a user will spread privacy information after receiving it. It is also important to determine to which receivers to send the information, making the game suitable for the strategic choice of both users in a specific scenario. The model should microscopically express the user's psychology and can thus be used to determine to which neighboring users the privacy information should be sent. Consequently, this paper applies the game theory to develop a model to facilitate the study of the dissemination of privacy information in OSNs based on the influencing factors of relationships between users and analyzes the main factors that would affect the dissemination of privacy information in OSNs from different perspectives.

3. A Game-Based Model for Privacy Information Dissemination

There are two major issues that need to be resolved in the construction of the model. The first is that after a user receives the privacy information, he/she needs to make a decision on whether to continue disseminating the information and on what probability to continue disseminating the information. The second is that after the user makes the decision to disseminate the privacy information, he/she needs to decide which neighboring user(s) to forward the privacy information. Based on the user's microscopic point of view, the proposed model would establish the user's sending probability and the user's preference probability, which are jointly affected by the two parameters of intimacy and popularity, and determine the forwarding probability to solve the first problem. The second problem is then solved by constructing a sender–receiver game model where the sender and each receiver play a pairwise game to determine the probability of each receiver's preference and the result of the Nash equilibrium threshold to determine whether the sender will send the privacy information. We assume that the dissemination of privacy information can be initiated by any node and that the privacy information has been obtained by the initiating node. We also assume that the privacy-concerning subject can be known by analyzing the content of the information. Privacy information dissemination in the proposed model can be schematically described as follows after a node obtains or receives a piece of privacy information.

(1) Check to see whether the information has already been received and processed. If yes, terminate the dissemination; if not, proceed to the next step.

(2) Calculate the probability of sending the information.

(3) Get the list of neighbor nodes, which are the potential receiving nodes.

(4) For each receiving node in the list, determine whether the information will be sent to the receiving node. If yes, the information is sent to the receiving node and the receiving

node will become a sending node in the next round of dissemination; if not, terminate the dissemination.

(5) Repeat steps (1)–(4) until there is no more sending node, which would finish the dissemination of the privacy information.

The complete process after a sender receives the privacy information is shown in Figure 1.

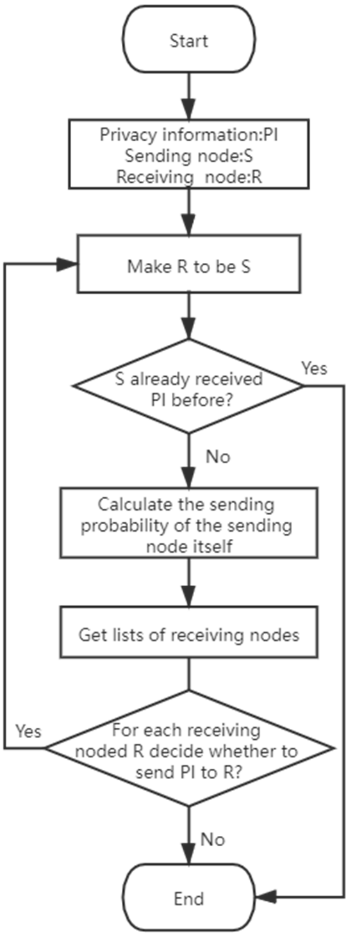
Start
Privacy information:PI
Sending node:S
Receiving node:R
Make R to be S
S already received PI before?
Yes
No
Calculate the sending probability of the sending node itself
Get lists of receiving nodes
For each receiving noded R decide whether to send PI to R?
Yes
No
End

Figure 1. The procedure of privacy information dissemination.

3.1. Decision on Forwarding Privacy Information

Considering the importance of the micro-relationship between users on the spread of privacy information, our model incorporates two parameters: intimacy and popularity, to quantify the sender's sending probability and the receiver's preference probability and to be combined with a game theory model based on the receiver's Nash equilibrium probability of the sending strategy of the sender.

3.1.1. Intimacy

Intimacy is expressed as the degree of closeness between two users in an OSN. Studies have shown that intimacy plays a vital role in terms of with which users would share their privacy information. 95.24% of people will choose whether to communicate personal privacy information based on perceived intimacy. The most intuitive example is that users with a high degree of intimacy, such as parents and children, would tend to protect the privacy of each other, i.e., to stop the spreading of privacy information. On the other hand, users with a low degree of intimacy, such as ordinary colleagues or even strangers, would not care much about the protection of each other's privacy and would, thus, have the tendency of spreading each other's privacy information. These examples indicate that intimacy has a great influence on whether a user will be interested and participate in the dissemination of privacy information.

Intimacy is defined as the emotional experience when an individual gets along well with others according to "The Dictionary of Psychology". It includes, but is not limited to, the affection between parents, brothers, and sisters, the love between men and women, the friendship between friends, and other different emotions. To facilitate our study, we classify intimacy into four levels in increasing order as shown in Table 1.

Table 1. Classification of intimacy levels.

Degree of Intimacy	Example
Family relationship	Parents, sisters, and other family members
Friend relationship	Friends, classmates, and roommates
Acquaintance relationship	Alumni, colleagues, and neighbors
Stranger relationship	Passersby in life, and strangers

3.1.2. Popularity

In OSNs, compared to the real social world, it is easier for someone to get the privacy of public figures and to participate in spreading it to others. The privacy exposure of public figures accounts for approximately 70% of the total privacy exposure on the Internet. On the contrary, the privacy information of ordinary people with a tiny number of fans will not receive widespread attention and people tend not to spread such privacy information. Therefore, dissemination of privacy information through examining popularity cannot be simply neglected.

The degree of a node in the network is the number of connections or edges between the nodes. In an OSN, the greater the degree of a user node, the more neighbor nodes the user has, and the higher the probability that the node is known in the network, the higher the popularity. The popularity of a node can be measured using Equation (1) in which degree(i) represents the degree of the current node, i.e., the number of fans in the network, and max degree is the largest degree of any node in the network.

$$P_{i\_famous} = \frac{\log_{10} \mathrm{degree(i)}}{\log_{10} degree(maxdegree)} \tag{1}$$

The popularity distribution of nodes in the network can be obtained through calculation. Through repeated experiments, we can get the areas of the distribution of popularity with three different intervals. Based on the intervals, the popularity of the privacy subjects can be classified into the following three kinds:

(1) High popularity users: This type of users each has a large number of fans. They are generally highly known and followed by a lot of people on the social network, such as celebrities and high-level politicians.

(2) Middle popularity users: This type of users each is usually well-known in a certain field or topic range of the social networks with a reasonable number of fans and followers in a specific field, such as food bloggers and game anchors.

(3) Low popularity users. This type of users are normally ordinary people with low levels of activity and only a few fans. Because of their low popularity, generally speaking, they do not attract much attention from other users.

3.1.3. Probability of Sending

The behavior of the sender is usually influenced by the degree of intimacy with the privacy subject and the popularity of the privacy subject. Then, sender vi can estimate its probability of forwarding the privacy information to other users based on the intimacy with the privacy subject and the popularity of the privacy subject, which is expressed in Formula (2), where $P_{intimacy\_is}$ represents the intimacy between the sending node and the privacy subject, P_{famous_s} represents the popularity of the privacy subject, and P_{send} will be used to determine the user's forwarding behavior of the private information. The receiver also measures its interest in the privacy message based on the intimacy with the privacy subject and the popularity of the privacy subject, which is expressed in Equation (3), where P_{intacy_js} represents the intimacy between the receiving node and the privacy subject and P_{like_js} is used to determine how much the receiver would like the privacy information.

$$P_{send} = (1 - P_{intimacy\_is})^{\gamma} \times P_{famous\_s} \tag{2}$$

$$\begin{cases} P_{like_{js}} = \alpha \sqrt[\eta]{P_{intimacy\_js}} + \beta \sqrt[\theta]{P_{famous\_s}} \\ \alpha + \beta = 1 \\ \eta, \theta > 1 \end{cases} \tag{3}$$

3.2. Decision on the Recipients of Privacy Information

The decision by sending a node to select a neighbor node is modeled using a game. At the micro level of the network structure, the game process is based on the interests of the nodes participating in the game. The whole process describes the strategy choices that the sending node sends to the receiver when the sending node has the privacy message. As shown in Table 2, at the micro level of the network structure, for a sending node and a receiving node, the policies for both parties are "send privacy information" and "do not send privacy information", respectively, and "like current privacy information" and "dislike current privacy information", respectively, which depicts the possibility that the node would send the private messages. In the following, we use the game theory to analyze the strategy choices of both the sender and the receiver in the game.

Table 2. The sender–receiver game matrix based on game theory.

	Receiver	
Sender	Like	Does not like
Send	$F_{-income}\begin{matrix} like \\ send \end{matrix}, J_{-income}\begin{matrix} like \\ send \end{matrix}$	$-F_{\_loss}\begin{matrix} dislike \\ send \end{matrix}, -J_{-loss}\begin{matrix} dislike \\ send \end{matrix}$
Does not send	$-F_{-loss}\begin{matrix} like \\ dont \end{matrix}, -J_{-loss}\begin{matrix} like \\ dont \end{matrix}$	$F_{\_income}\begin{matrix} dislike \\ dont \end{matrix}, J_{-income}\begin{matrix} dislike \\ dont \end{matrix}$

First, we need to define the corresponding benefit when different strategies are adopted by the two sides of the game:

$F_{-income}\begin{matrix} like \\ send \end{matrix}$: This is the benefit to the sender when the receiver adopts strategy "like privacy information" and the sender adopts strategy "send privacy information". The benefit can be viewed in the way that the sender gains more attention from the receiver by sending the privacy information, potentially making the relationship between the two parties closer.

$J_{-income}\begin{matrix} like \\ send \end{matrix}$: This is the benefit to the receiver when the receiver adopts strategy "like privacy information" and the sender adopts strategy "send privacy information". The benefit can be viewed in the way that the receiver gets information to satisfy curiosity and pleasure.

$-F_{\_loss}\begin{matrix} dislike \\ send \end{matrix}$: This is the loss to the sender when the receiver adopts strategy "do not like privacy information" and the sender adopts strategy "send privacy information". The loss can be viewed in the way that the disclosure of some other's privacy information by the sender causes some discomfort, which could result in bad feelings that may lead to the loss of attention or the consequence of some harm to the sender.

$-J_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send}$: This is the loss to the receiver when the receiver adopts strategy “do not like privacy information” and the sender adopts strategy “send privacy information”. The loss can be viewed in the way that the receiver may generate some sense of dissatisfaction and disgust when receiving a message that he/she does not like.

$-F_{\_loss}\genfrac{}{}{0pt}{}{like}{dont}$: This is the loss to the sender when the receiver adopts strategy “like privacy information” and the sender adopts strategy “do not send privacy information”. The loss can be viewed in the way that the sender loses an opportunity to gain more attention from the receiver.

$-J_{\_loss}\genfrac{}{}{0pt}{}{like}{dont}$: This is the loss to the receiver when the receiver adopts “like privacy information” and the sender adopts strategy “do not send privacy information”. The loss can be viewed in the way that the receiver misses receiving a message that he/she likes.

$F_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont}$: This is the benefit to the sender when the receiver adopts strategy “do not like privacy information” and the sender adopts strategy “do not send privacy information”. The benefit can be viewed in the way that not sending a message that the receiver does not like can continue maintaining the receiver’s attention to the sender.

$J_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont}$: This is the benefit to the receiver when the receiver adopts strategy “do not like privacy information” and the sender adopts strategy “do not send privacy information. The benefit can be viewed in the way that the receiver does not receive the message that he/she does not like.

Now, we need to analyze the game matrix in Table 2. From the perspective of the receiver, when the sender chooses the strategy “send privacy information “, adopting the strategy “like privacy information” by the receiver can bring greater benefits, i.e., $J_{\_income}\genfrac{}{}{0pt}{}{like}{send} > -J_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send}$. When the sender chooses strategy “do not send privacy information”, adopting strategy “do not like privacy information” by the receiver can bring greater benefits, i.e., $J_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont} > -J_{\_loss}\genfrac{}{}{0pt}{}{like}{send}$. From the perspective of the sender, when the receiver chooses strategy “like privacy information”, adopting strategy “send privacy information” by the sender can bring greater benefits, i.e., $F_{\_income}\genfrac{}{}{0pt}{}{like}{send} > -F_{\_loss}\genfrac{}{}{0pt}{}{like}{dont}$. When the receiver chooses strategy “do not like privacy information”, adopting strategy “send privacy information” brings less benefits than adopting strategy “do not send privacy information” by the sender, i.e., $-F_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send} < F_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont}$. From the above analysis, we can see that there is no pure strategy Nash equilibrium in the game matrix, so we need to calculate its mixed strategy Nash equilibrium.

P_F and P_J is set as the income matrix of the sender and the receiver, respectively. Assuming that the probability that the sender chooses strategy “send privacy information” is x, then the probability that the sender chooses strategy “do not send privacy information” is 1 − x and the mixed strategy probability of the sender is P_f = (x,1 − x). Similarly, assuming that the probability that the receiver chooses strategy “like privacy information” is y, then the probability that the receiver chooses strategy “do not like privacy information” is 1 − y and the mixed strategy probability of the receiver is P_j = (y,1 − y). The sender’s revenue function can be calculated by Equation (4):

$$\begin{aligned}\mathrm{E_F} &= P_f \times P_F \times P_j^T \\ &= [x \;\; 1-x]\begin{bmatrix} F_{\_income}\genfrac{}{}{0pt}{}{like}{send} & -F_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send} \\ -F_{\_loss}\genfrac{}{}{0pt}{}{like}{dont} & F_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont} \end{bmatrix}\begin{bmatrix} y \\ 1-y \end{bmatrix} \\ &= x \times y \times F_{\_income}\genfrac{}{}{0pt}{}{like}{send} + x \times (1-y) \times \left(-F_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send}\right) + \\ &\quad (1-x) \times y \times \left(-F_{\_loss}\genfrac{}{}{0pt}{}{like}{dont}\right) + (1-x) \times (1-y) \times F_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont}\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{\alpha E_F}{\alpha x} &= y \times F_{\_income}\genfrac{}{}{0pt}{}{like}{send} + (1-y) \times \left(-F_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send}\right) + y \times F_{\_loss}\genfrac{}{}{0pt}{}{like}{dont} + \\ &\quad (y-1) \times F_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont}\end{aligned} \tag{5}$$

By forcing Equation (5) to be equal to 0, we can obtain the value of y, which can be expressed using Equation (6):

$$y = \frac{F_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont} + F_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send}}{F_{\_income}\genfrac{}{}{0pt}{}{like}{send} + F_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send} + F_{\_loss}\genfrac{}{}{0pt}{}{like}{dont} + F_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont}} \tag{6}$$

Similarly, the receiver's income P_J can be calculated using Equation (7):

$$\begin{aligned}\mathrm{E_J} &= P_f \times P_J \times P_j^T \\ &= [x \;\; 1-x]\begin{bmatrix} J_{\_income}\genfrac{}{}{0pt}{}{like}{send} & -J_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send} \\ -J_{\_loss}\genfrac{}{}{0pt}{}{like}{dont} & J_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont} \end{bmatrix}\begin{bmatrix} y \\ 1-y \end{bmatrix} \\ &= x \times y \times J_{\_income}\genfrac{}{}{0pt}{}{like}{send} + x \times (1-y) \times \left(-J_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send}\right) + \\ &\quad (1-x) \times y \times \left(-J_{\_loss}\genfrac{}{}{0pt}{}{like}{dont}\right) + (1-x) \times (1-y) \times J_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont}\end{aligned} \tag{7}$$

Taking the derivative of y, we can get Equation (8):

$$\begin{aligned}\frac{\alpha E_J}{\alpha y} &= x \times J_{\_income}\genfrac{}{}{0pt}{}{like}{send} + x \times J_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send} + (1-x) \times \left(-J_{\_loss}\genfrac{}{}{0pt}{}{like}{dont}\right) + \\ &\quad (x-1) \times J_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont}\end{aligned} \tag{8}$$

Forcing Equation (8) to be equal to 0, we can obtain the value of x, which can be expressed using Equation (9):

$$x = \frac{J_{\_income}\genfrac{}{}{0pt}{}{like}{send} + J_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send}}{J_{\_income}\genfrac{}{}{0pt}{}{like}{send} + J_{\_loss}\genfrac{}{}{0pt}{}{dislike}{send} + J_{\_loss}\genfrac{}{}{0pt}{}{like}{dont} + J_{\_income}\genfrac{}{}{0pt}{}{dislike}{dont}} \tag{9}$$

The resulting mixed strategy Nash equilibrium is then shown in Equations (10) and (11):

$$\begin{aligned}&[x \quad 1-x] \\ &= \left[\frac{J_{\ income}\begin{matrix} like \\ send \end{matrix} + J_{\ loss}\begin{matrix} dislike \\ send \end{matrix}}{J_{\ income}\begin{matrix} like \\ send \end{matrix} + J_{\ loss}\begin{matrix} dislike \\ send \end{matrix} + J_{\ loss}\begin{matrix} like \\ dont \end{matrix} + J_{\ income}\begin{matrix} dislike \\ dont \end{matrix}} 1 \right. \\ &\left. - \frac{J_{-income}\begin{matrix} like \\ send \end{matrix} + J_{\_loss}\begin{matrix} dislike \\ send \end{matrix}}{J_{-income}\begin{matrix} like \\ send \end{matrix} + J_{\_loss}\begin{matrix} dislike \\ send \end{matrix} + J_{-loss}\begin{matrix} like \\ dont \end{matrix} + J_{\_income}\begin{matrix} dislike \\ dont \end{matrix}} \right]\end{aligned} \tag{10}$$

$$\begin{aligned}&[y \quad 1-y] \\ &= \left[\frac{F_{\_income}\begin{matrix} dislike \\ dont \end{matrix} + F_{\_loss}\begin{matrix} dislike \\ send \end{matrix}}{F_{-income}\begin{matrix} like \\ send \end{matrix} + F_{\_loss}\begin{matrix} dislike \\ send \end{matrix} + F_{-loss}\begin{matrix} like \\ dont \end{matrix} + F_{\_income}\begin{matrix} dislike \\ dont \end{matrix}} 1 \right. \\ &\left. - \frac{F_{\_income}\begin{matrix} dislike \\ dont \end{matrix} + F_{\_loss}\begin{matrix} dislike \\ send \end{matrix}}{F_{-income}\begin{matrix} like \\ send \end{matrix} + F_{\_loss}\begin{matrix} dislike \\ send \end{matrix} + F_{-loss}\begin{matrix} like \\ dont \end{matrix} + F_{\_income}\begin{matrix} dislike \\ dont \end{matrix}} \right]\end{aligned} \tag{11}$$

Namely:

$$P_{like\_nash} = \frac{J_{-income}\begin{matrix} like \\ send \end{matrix} + J_{\_loss}\begin{matrix} dislike \\ send \end{matrix}}{J_{-income}\begin{matrix} like \\ send \end{matrix} + J_{\_loss}\begin{matrix} dislike \\ send \end{matrix} + J_{-loss}\begin{matrix} like \\ dont \end{matrix} + J_{\_income}\begin{matrix} dislike \\ dont \end{matrix}} \tag{12}$$

By deriving the above mixed strategy Nash equilibrium, the probability that the sender chooses strategy "sending privacy information" and the probability that the receiver chooses strategy "like privacy information" can be obtained. We can also see that the Nash equilibrium point of the receiver's preference probability is P_{like_nash} as shown in Equation (12). If the probability that the receiver likes privacy information is lower than the threshold, the sender should not send the information to get more benefit, i.e., the sender should not send the privacy information to the receiver. On the other hand, if the probability that the receiver likes privacy information is higher than the threshold, the sender will send the information to get more benefit, i.e., the sender will send the privacy information to the receiver.

4. Experiment and Result Analysis

To study the effect of intimacy, popularity and the Nash equilibrium threshold on the model, we first used the Facebook's network structure data that we obtained with N = 4039 nodes or from users to construct a social network in which we set P_{like_nash} = α = β = 0.5, η = 2, θ = 3, and γ = 2. As can be seen in Figure 2, the network follows the power-law distribution, which conforms to the basic characteristics of the scale-free network, namely the BA network. We also used an algorithm to build a BA scale-free network with N = 4500 nodes, which is comparable to the Facebook network in terms of scale. During the construction, the initial number of users was set to be m_0 = 5 and the number of connections that was added when a new user joined the network was m_1 = 5. The degree distribution of the BA scale-free network thus constructed is depicted in Figure 3. We can thus see from Figures 3 and 4 that the two networks with one being based on real Facebook data and the other being constructed for the experiment have a similar network structure in terms of degree distribution.

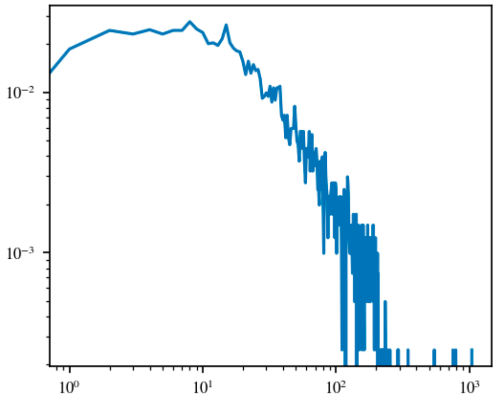

Figure 2. Degree distribution in the Facebook network.

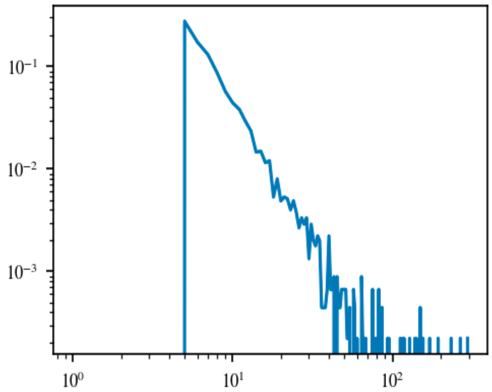

Figure 3. Degree distribution in the constructed BA network.

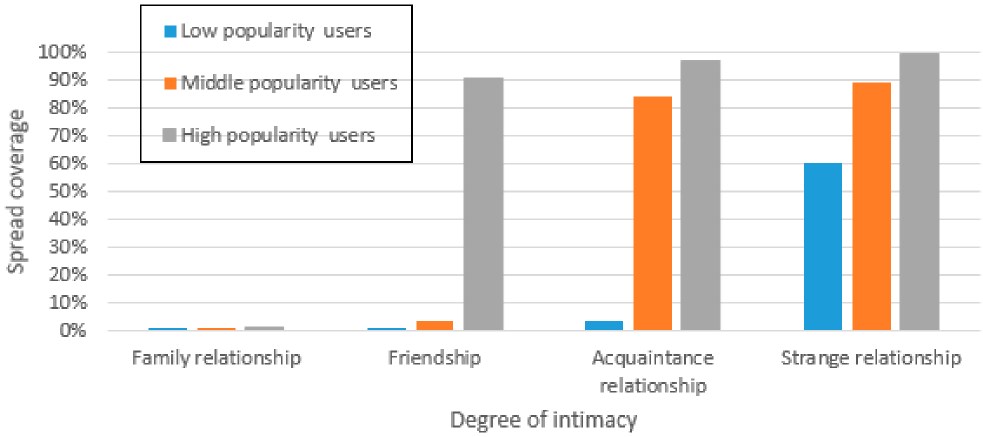

Figure 4. Range of privacy dissemination vs. intimacy and popularity.

Our experiment aims to both evaluate and validate the proposed privacy information dissemination model. The strategy of the experiment is to use the constructed BA scale-free network to evaluate our model through examining how privacy information is disseminated in the network and the Facebook network data to validate our model through comparing the results of the BA scale-free network and the Facebook network.

In the experiment, we divided intimacy into four levels: family affection, friendship, acquaintance, and stranger relationship, whose intervals and the proportion of users are shown in Table 3.

Table 3. Proportion of intimacy and distribution of value range.

Level of Intimacy	Intimacy Threshold	Proportion of Users
Family relationship	[0.8,1)	1%
Friendship	[0.5,0.8)	4%
Acquaintance relationship	[0.2,0.5)	5%
Stranger relationship	[0,0.2)	90%

In our experiment, we divided popularity into three levels: high popularity, middle popularity, and low popularity. According to formula (1), we can calculate each user's popularity in the network. We found after repeated experiment that the average value of the popularity of three types of users is around 0.8, 0.5, and 0.3, respectively. Considering that in a real situation, the degree of awareness of a user to a particular user is not the same, the popularity of other users in the network is set according to the normal distribution of their popularity to better reflect the reality as shown in Table 4.

Table 4. Proportion of popularity and distribution of value range.

Types of Popularity	Popularity Value	Proportion
High popularity users	$C \sim N(0.8,0.2)$	1%
Middle popularity users	$C \sim N(0.5,0.2)$	20%
Low popularity users	$C \sim N(0.3,0.2)$	79%

4.1. Results and Analysis

Figure 4 shows the range of privacy information dissemination for the four different levels of intimacy and three levels of popularity. When a family member serves as the sending node, regardless of the popularity of the privacy subject, the tendency is to protect the privacy of the subject. When a friend of the privacy subject receives a private message, he/she would only choose to disseminate the private message of a high popularity user who is sensational and attractive but block other types of dissemination. Acquaintances would block the dissemination of privacy information with low sensation and attractiveness, while strangers would not exercise much protection on the privacy of the subject.

Since popularity affects the desire of social network users to disseminate the privacy information of the subject, we now explore the influence of the types of privacy subjects on the dissemination of privacy information, and the results are shown in Figures 5–7. As can be seen, in the BA network, the spread of privacy information of high popularity users progresses much faster, taking just a few rounds to cover almost all the users in the network before the dissemination ends. For middle popularity users, because they maintain a certain degree of popularity and attractiveness in the social network, their privacy information will be disseminated in the network at a relatively stable rate, and their privacy information can still be acquired by a lot of users to achieve greater popularity although the total number of users is approximately 30% less. For low popularity users, their privacy information spreads in the network at a much slower rate but may last for more rounds. Ultimately, their privacy information would still be received by a small percentage of users.

In addition, the Nash equilibrium threshold of the receiver's preference for privacy information also impacts the dissemination. The lower the threshold, i.e., the greater the probability that the receiver likes it, the higher the probability that the sender chooses to send. As shown in Figure 8, as the receiver's Nash equilibrium threshold for privacy information gets lower, the receiver becomes more likely to like the privacy information. The sender would then choose to send privacy information to get more benefits, making the sender more likely to send the privacy information. When the Nash equilibrium threshold becomes higher, the receiver is more likely to dislike the message. So, in order not to arouse the receiver's disgust, the sender would choose not to send it to maintain the receiver's continuous attention. When the Nash equilibrium threshold becomes extremely high, the sender is afraid of spreading and most of users choose not to send others' privacy, resulting in the privacy information to be well protected.

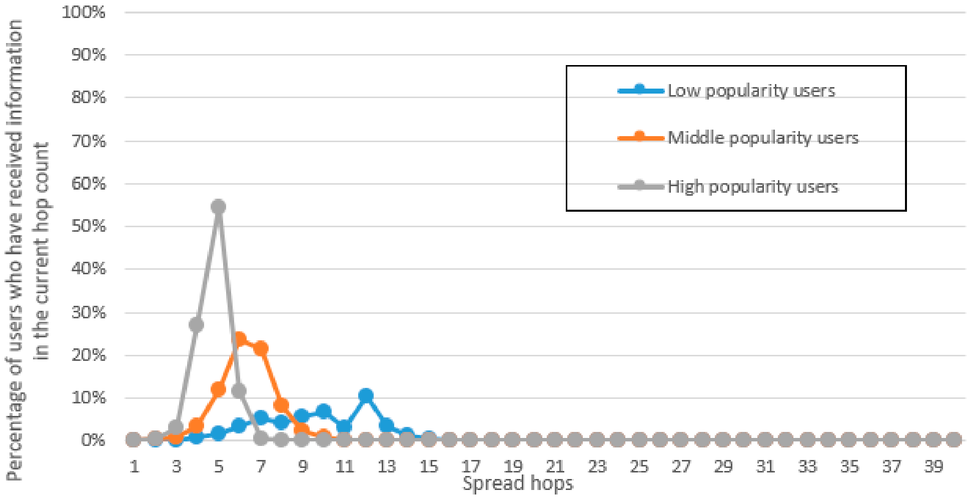

Figure 5. Privacy dissemination in each round vs. popularity.

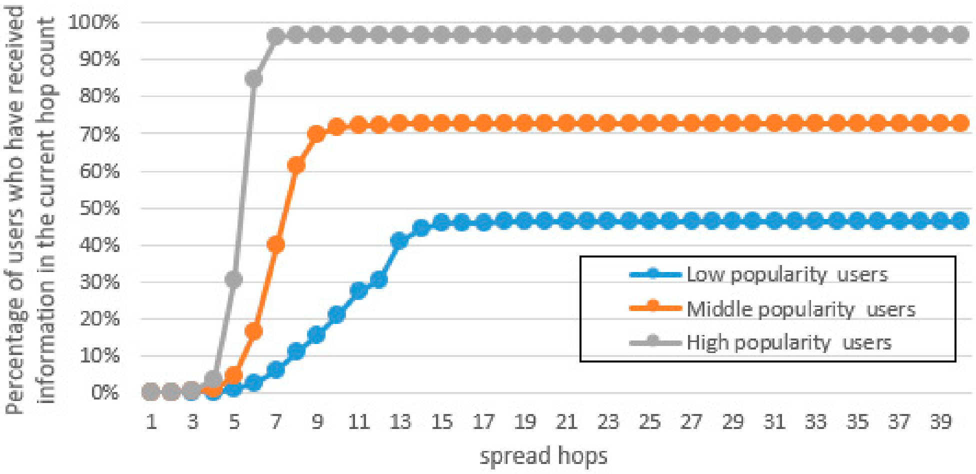

Figure 6. Range of privacy dissemination vs. popularity.

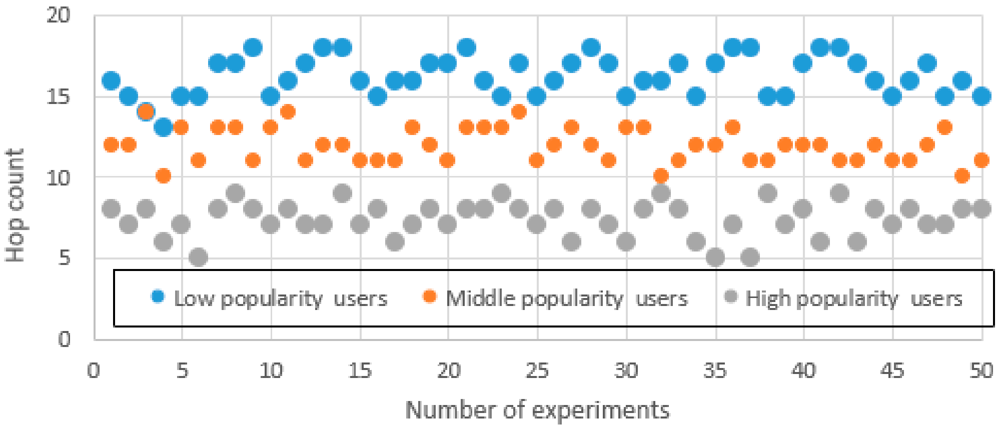

Figure 7. Results in different experiments for different levels of popularity.

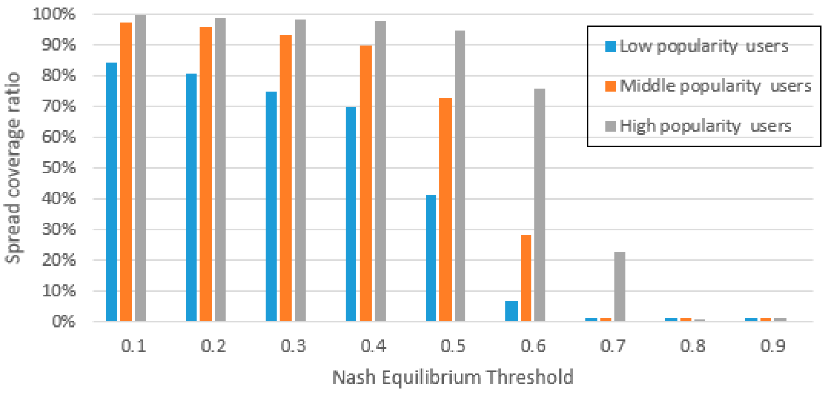

Figure 8. Range of dissemination vs. popularity under different Nash equilibrium thresholds.

4.2. Validation of the Model

Figure 9 shows the dissemination of privacy information in the BA network and in the Facebook network. As the Nash equilibrium continues to improve, the two models display the same downward trend and the final coverage is almost the same.

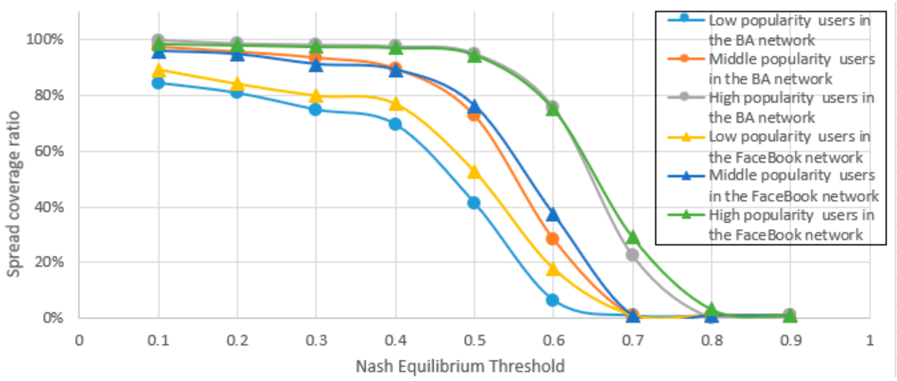

Figure 9. Range of privacy dissemination under different Nash equilibrium thresholds in the BA and the Facebook networks.

When categorizing the popularity of privacy subjects as high popularity users, meddle popularity users, and low popularity users, the current number of hops in the BA network and Facebook network to receive information and the total number of users receiving the information change with the number of hops are shown in Figures 10 and 11, respectively.

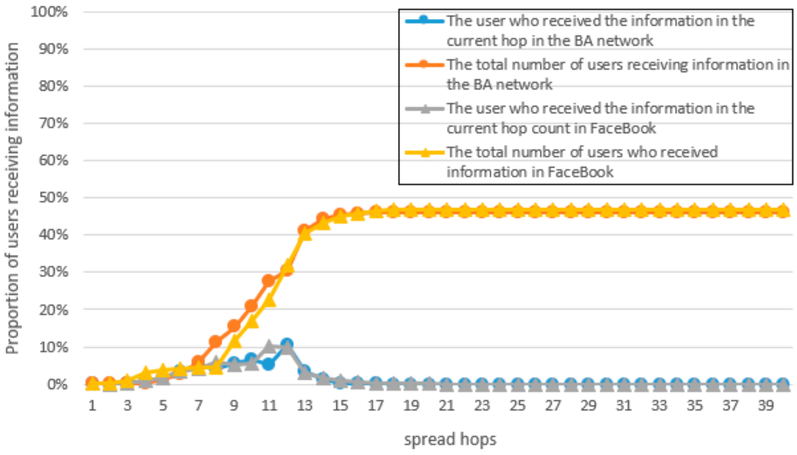

Figure 10. The dissemination process of the privacy information of high popularity users in the BA and the Facebook networks.

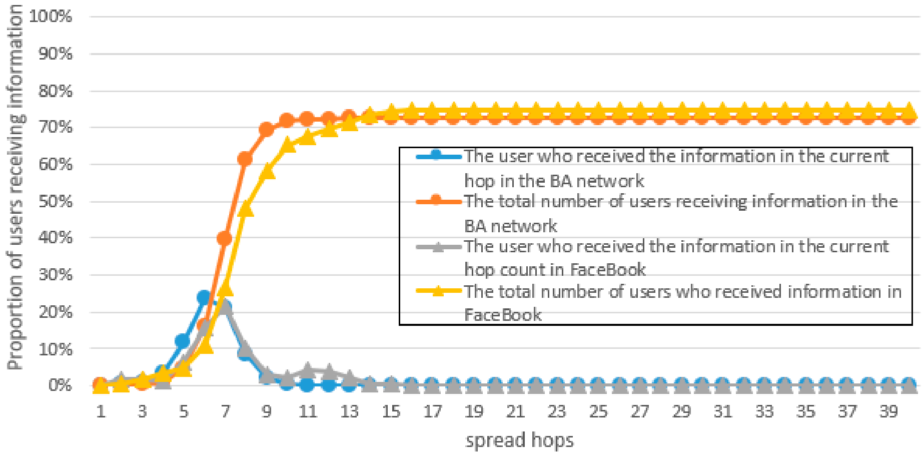

Figure 11. The dissemination process of the privacy information of middle popularity users on the BA network and the Facebook network.

Figure 12 shows that the final coverage, the speed and the dissemination trend are highly consistent. Thus, our proposed model can be used to simulate the process of privacy information dissemination in OSNs such as Facebook reasonably and effectively.

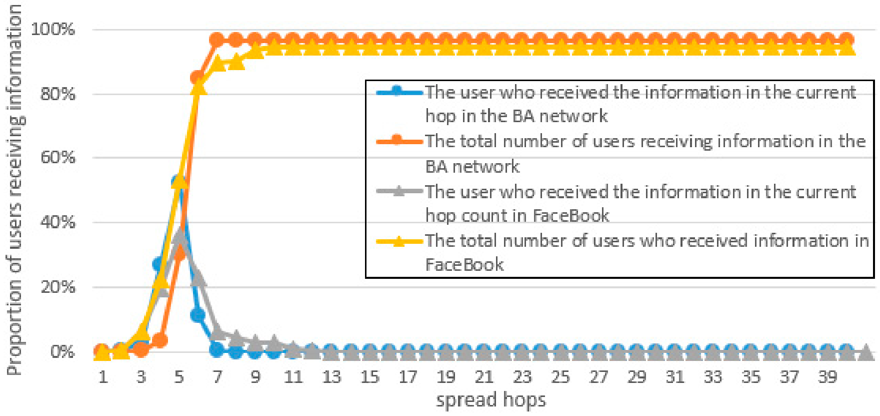

Figure 12. The dissemination process of the privacy information of low popularity users on the BA network and the Facebook network.

Additional model verification was done based on the classification of users in which the Xinbang data platform (https://data.newrank.cn/) was crawled to get the data for the experiment. The Xinbang platform focuses on analyzing data and predicting future trends based on data provided by third-party platforms that have a large number of users, such as Tencent's WeChat, ByteDance's Douyin (TikTok China), etc. It analyzes business and user data and predicts future trends for industries and enterprises by collaborating with large companies. Powered with proprietary analytical software, the Xinbang platform has been one of the major prediction platforms in China for many years with the partnership with WeChat and Douyin that have accumulated 1.2 billion and 920 million users, respectively. In the experiment, Xinbang was searched using several keywords to obtain the privacy information of three types of users as well as the spreading of the privacy information of these users during the period of dissemination. The keywords used in the search for the three different types of users are shown in Table 5.

Table 5. Types of selected users and search keywords.

Types of Popularity	Search Keyword
High popularity users	Star A sex scandal
Middle popularity users	Anchor B was bursting in love
Low popularity users	Husband C Domestic Violence Wife

As shown in Figures 13–15, for user with different level of popularity, the model's simulation data and the real statistical data are highly consistent in terms of the change and the overall trend. Although there is a slight difference in terms of the time that it takes to reach the peak, the overall performance is similar, which can be used to simulate the privacy information dissemination process of different types of users in OSNs. These results suggest that the model proposed in this paper can largely reflect the dissemination of privacy information of various privacy subjects in OSNs and the experimental results can provide a valuable reference for understanding the dissemination of privacy information by social network users.

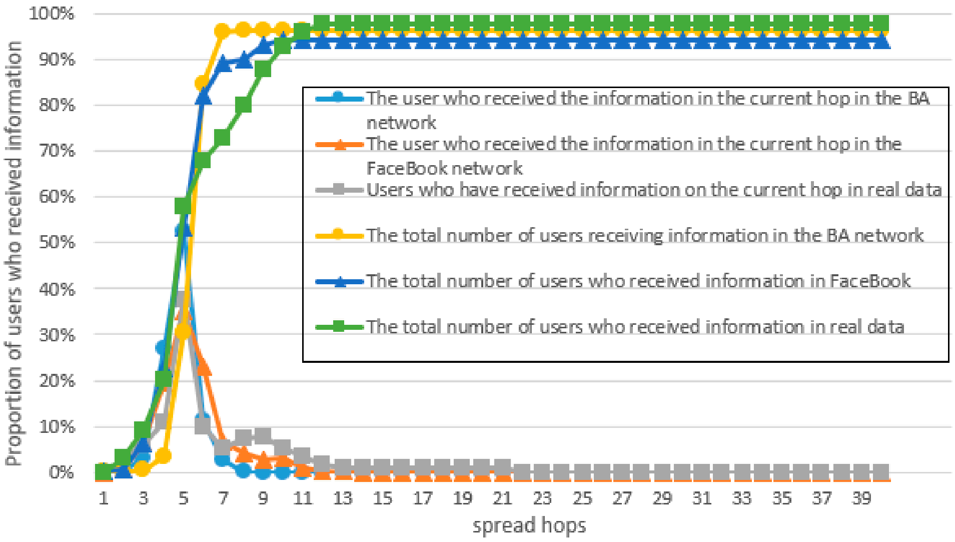

Figure 13. Dissemination of the privacy information of a highly popular user between simulated data and real data.

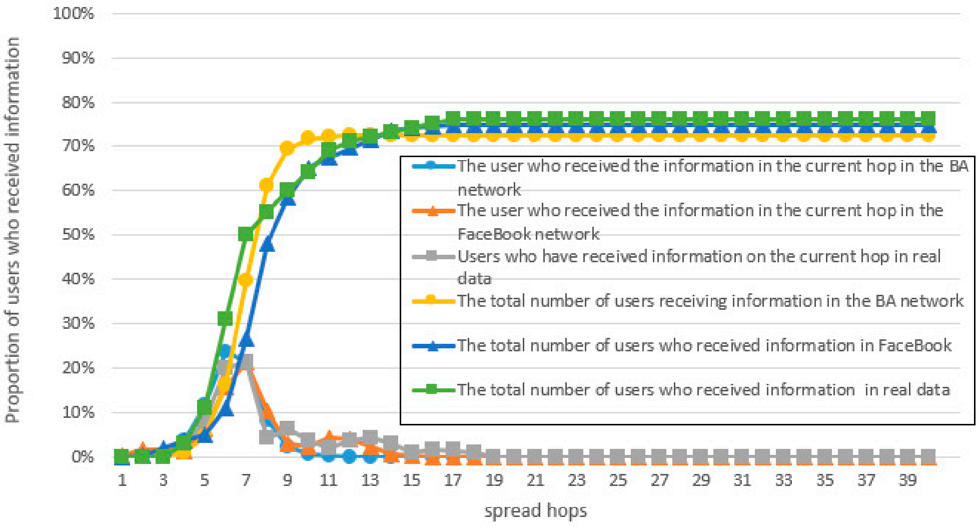

Figure 14. Dissemination of the privacy information of a medium popular user between simulated data and real data.

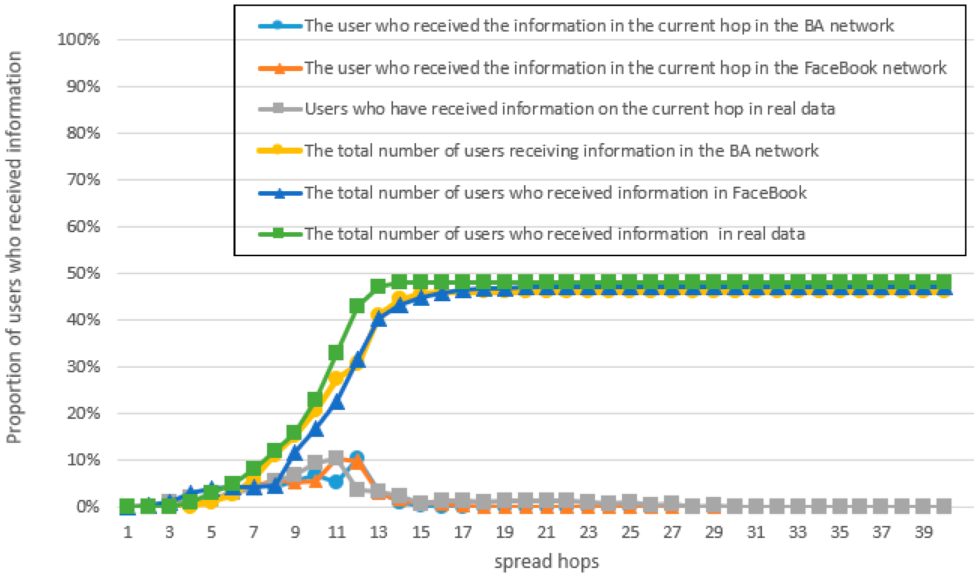

Figure 15. Dissemination of the privacy information of a low popular user between simulated data and real data.

The following conclusions can be drawn after comparison to the real data: the higher the intimacy between the initial communication node and the privacy subject, the more likely the receiver will protect the privacy and block the spread of privacy information. In contrast, the lower the intimacy, the easier it is to spread the privacy information. Users tend to disseminate the privacy information of users with high popularity. The scope of privacy information dissemination of high popularity users is much larger than that of low popularity users. The higher the popularity, the faster the information dissemination, and it can be spread throughout the whole only after a few rounds of forwarding net. The smaller the Nash equilibrium probability that the receiver likes, the greater the probability that the receiver likes it, the greater the possibility that the sender chooses to send, and the wider the coverage of information.

5. Conclusions

This paper firstly introduced the current stage of research on privacy information dissemination in online social networks by analyzing their characteristics and limitations. Secondly, the strategy choices of the users during the dissemination of privacy information were analyzed by applying the game theory and taking the profits of the users into consideration. On this basis, this paper introduced two factors that would affect the dissemination of privacy information, i.e., intimacy and attention, and proposed a game theory-based privacy information dissemination model that could be used to simulate the dissemination process of privacy information. This paper subsequently described the implementation process, player interaction process, and specific architecture design of the proposed model based on game theory and compared it to real social networks to validate the model through repeated experiments. The experimental results showed that the proposed privacy information dissemination model conforms well to the dissemination of privacy information in OSNs and can thus be used to conduct study aiming at understanding the general pattern or trend of privacy information dissemination in OSNs.

In the future, we will continue improving the proposed privacy information dissemination model. One direction of the improvement could be on classifying the OSN users with finer granularities on both the intimacy and the popularity. Another possible direction of the improvement could be on considering the background information in the setup of the thresholds for intimacy and popularity to make them more dynamic and more suitable for real OSNs. The improvement will make our privacy information dissemination model capable of adapting to OSNs of different types and various sizes.

Author Contributions: Conceptualization, J.H.; methodology, J.H., Y.L. and N.Z.; software, Y.L.; validation, Y.L. and N.Z.; formal analysis, N.Z.; investigation, Y.L.; resources, N.Z.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, J.H.; visualization, Y.L.; supervision, N.Z.; project administration, N.Z.; funding acquisition, N.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 61602456.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Review of the development of China's Internet industry in 2020. *Internet World* **2021**, *2021*, 16–19.
2. We Are Social. In *Global Digital 2021 Reports*; We Are Social Inc.: New York, NY, USA, 2021.
3. Kim, H.; Hovav, A.; Han, J. Protecting intellectual property from insider threats: A management information security intelligence perspective. *J. Intellect. Cap.* **2019**, *21*, 181–202. [CrossRef]
4. Zhou, M.; He, H.; Fu, Z.; Zhuo, Z. Role extraction in complex networks and its application in control of networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *44*, 246–248. [CrossRef]
5. Watts, D.J.; Strogatz, S.H. Collective dynamics of "small-world" networks. *Nature* **1998**, *393*, 440–442. [PubMed]
6. Barabasi, A.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [CrossRef] [PubMed]
7. Osborne, M.; Rubinstein, A. *A Course in Game Theory*; The MIT Press: Cambridge, MA, USA, 1994.
8. Von Neumann, J.; Morgenstern, O. *Theory of Games and Economic Behavior*; Princeton University Press: Princeton, NJ, USA, 1944.
9. Nash, J.F. Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 48–49. [CrossRef] [PubMed]
10. Chen, X.; Van Der Lans, R.; Trusov, M. Efficient estimation of network games of incomplete information: Application to large online social networks. *Manag. Sci.* **2021**, *67*, 7575–7598. [CrossRef]
11. Weyrich, P.; Ruin, I.; Terti, G.; Scolobig, A. Using serious games to evaluate the potential of social media information in early warning disaster management. *Int. J. Disaster Risk Reduct.* **2021**, *56*, 10253. [CrossRef]
12. Gorelov, M.A. Topological statement of the information aggregation problem in hierarchical games. *Autom. Remote Control.* **2021**, *82*, 308–323. [CrossRef]
13. Simon, R.; Spież, S.; Toruńczyk, H. Games of incomplete information and myopic equilibria. *Isr. J. Math.* **2017**, *241*, 721–748. [CrossRef]
14. Huang, H.; Wang, T.; Hu, M.; Dong, M.; Lai, L. Node attitude aware information dissemination model based on evolutionary game in social networks. *Mob. Netw. Appl.* **2021**, *26*, 114–129. [CrossRef]
15. Lai, W.K.; Chen, Y.U.; Wu, T.Y. Analysis and evaluation of random-based message propagation models on the social networks. *Comput. Netw.* **2020**, *170*, 107047. [CrossRef]
16. Hartmann, P.; Fernández, P.; Apaolaza, V.; Eisend, M.; D'Souza, C. Explaining viral CSR message propagation in social media: The role of normative influences. *J. Bus. Ethics* **2021**, *173*, 365–385. [CrossRef]
17. Gu, K.; Wang, L.; Yin, B. Social community detection and message propagation scheme based on personal willingness in social network. *Soft Comput.* **2019**, *23*, 6267–6285. [CrossRef]
18. Brusco, M.J.; Steinley, D.; Stevens, J.; Cradit, J.D. Affinity propagation: An exemplar-based tool for clustering in psychological research. *Br. J. Math. Stat. Psychol.* **2019**, *72*, 155–182. [CrossRef] [PubMed]
19. Bi, H. Aggregation encryption method of social network privacy data based on matrix decomposition algorithm. *Wirel. Pers. Commun.* **2022**, *127*, 369–383. [CrossRef]
20. Adjei, J.K.; Adams, S.; Mensah, I.K.; Tobbin, P.E.; Odei-Appiah, S. Digital identity management on social media: Exploring the factors that influence personal information disclosure on social media. *Sustainability* **2020**, *12*, 9994. [CrossRef]
21. Li, C.T.; Zeng, Z.Y. Learning effective feature representation against user privacy protection on social networks. *Appl. Sci.* **2020**, *10*, 4835. [CrossRef]
22. Bioglio, L.; Pensa, R. Modeling the impact of privacy on information diffusion in social networks. In Proceedings of the 8th Conference on Complex Networks, Dubrovnik, Croatia, 21–24 March 2017; pp. 95–107.
23. Bioglio, L.; Pensa, R. Impact of neighbors on the privacy of individuals in online social networks. *Procedia Comput. Sci.* **2017**, *108*, 28–37. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.