Personalized information diffusion in signed social networks

Cunquan Qu\textsuperscript{1,2,3,*}, Jialin Bi\textsuperscript{2,3} and Guanghui Wang\textsuperscript{1,2,3,*}

\textsuperscript{1} Data Science Institute, Shandong University, 250100, Jinan, People’s Republic of China
\textsuperscript{2} School of Mathematics, Shandong University, 250100, Jinan, People’s Republic of China
\textsuperscript{3} The authors declare no conflicts of interest.
* Authors to whom correspondence should be addressed.
E-mail: cqqu@sdu.edu.cn, jialinbi@mail.sdu.edu.cn and ghwang@sdu.edu.cn

Keywords: information diffusion, signed networks, structure balance, social networks

Abstract
Understanding the dynamics in complex networks is crucial in various applications, such as quelling the epidemic outbreak, preventing the spread of rumors online, and promoting the diffusion of science and technology information. In this study, we investigated a personalized information diffusion (PID) mechanism on signed networks. The main assumption of this mechanism is that if a message is good for the stakeholder, then it is also good for his/her friends but bad for his/her enemies. At each step, the individual who receives the information will determine whether to forward it based on his/her relationship with the stakeholder. We find that bad news may spread further than good news even if a stakeholder has more directly connected friends than enemies. Moreover, the nodes that have more (potential) friends across the network can spread good information more widely. However, individuals who have more enemies locally can spread bad information more widely. Our findings may inspire the design of strategies for controlling information, epidemics, or rumors in social networks.

1. Introduction
Understanding the mechanisms of information diffusion or spreading is one of the main objectives in network science, as it is essential in various fields, such as economics, biology, and social science [1–3]. The intrinsic complexities of the information and the environment render challenging the modeling and prediction of such processes. Many researchers conduct related studies under the assumption that the information’s intrinsic property might remain unchanged during the transmission, the relative attribute varies over time due to the heterogeneity of the nodes in the system [4].

Signed networks, which are specialized heterogeneous systems, have been widely investigated in the last several decades [5]. Nodes in signed networks are connected by two logically different types of links: positive and negative links [6]. Signed networks can well represent many real-world systems. The positive (negative) links may represent the promotion (inhibition) relationships in protein interaction networks. In a signed social network [7, 8], the positive and negative links represent friendly and antagonistic interactions, respectively. In the 1940s, Heider introduced the concept of signed networks and proposed the well-known structural balance theory in terms of triangles in social science [9]. According to structural balance theory triangles satisfy ‘the friend of my friend, as well as the enemy of my enemy, is my friend’ are balanced; otherwise, they are unbalanced. Later, Cartwright and Harary generalized structural balance theory to cases of longer cycles. According to the generalized theory, cycles with an odd number of negative links are unbalanced; otherwise, they are balanced [10].

With the development of complex networks and data science, there are growing concerns regarding topics in structural balance theory [11–13]. One of these topics is the design of algorithms for computing the structural balance of large-scale datasets, especially the cycle-based structural balance. In 2011, Facchetti et al generalized the algorithm for ground-state calculation in large scale Ising spin glasses to calculate the global structural balance, and showed that most large online social networks are incredibly balanced [7]. Recently,
Kirkley et al proposed two measures for weak and strong structural balance and further demonstrated the significant balance of real-world networks. The dynamic analysis of structural balance is also challenging, which explores how unbalanced structures become balanced [14]. In 2010, Marvel et al examined continuous-time models and concluded that the initial amount of friendliness determines whether the system will end up in a global balance [15].

Another topic regarding signed networks is dynamic analysis [16–18]. Most studies treat the disseminated information as fixed, and the information is of equal significance to every node in the system. Shi et al [17] investigated the evolution of opinions on signed networks. At each step, each node updates its state according to only the directly connected neighbors and the signs of the corresponding links. Wang et al [18] proposed the self-avoiding pruning random walk in signed networks, from which negative neighbors will be removed with a pre-specified probability. In the above models, every node’s attitude toward the opinion or walker is always the same. However, in real scenarios, the meaning of a message likely differs among people. In this study, we discuss the diffusion of personalized information on signed social networks. The model that is introduced here fully utilizes structural balance theory.

Changing, or evolution or variation from a biological perspective is crucial for transmitting information or an epidemic outbreak [19–21]. In many online communication platforms, spreading a message does not require changing or editing the message, such as spreading using the forward function in WeChat, Twitter, and Facebook. The intrinsic properties of the message remain unchanged, while the relative properties of the message differ among people. Here, we describe this spreading process as ‘receive without a difference and forward conditionally’. Once a message is published by a user, the followers obtain the information. However, whether a follower forwards the message depends on whether the follower is interested in it. We try to explore how this kind of information diffuses in the signed social network and give an explanation of a Chinese proverb ‘for evil news rides fast, while good news baits later’.

The remainder of the paper is organized as follows: in section 2, we introduce the basic definitions of signed networks and the proposed personalized information diffusion (PID)-process. In addition, we theoretically analyze three properties that are related to the proposed mechanism, the number of publishers, the ratio of publishers that publish the message, and the proportion of friends in the publisher group. We conduct experiments on synthetic networks in section 3. Our observations and understanding are examined on real-world signed networks in section 4. We discuss the results on real-world networks in section 5 and present the conclusions of this study in section 6.

2. The PID-mechanism for signed networks

In this section, we introduce basic notations of signed networks and the personalized (P) information (I) diffusion (D) mechanism on signed networks.

2.1. Notations and definitions

We consider an undirected signed network \( G = (V,E) \), where \( V = \{1, 2, \ldots, n\} \) is the node set and \( E \) is the edge set. Each edge in \( E \) is associated with a sign: positive or negative. Considering the signed social network as an example, \( V \) is the user set and \( E \) denotes the relationship set.

A signed network can be represented by its adjacency matrix \( A \). Hence,

\[
A_{ij} = \begin{cases} 
1, & \text{if } v_i \text{ and } v_j \text{ are connected via a positive edge}, \\
-1, & \text{if } v_i \text{ and } v_j \text{ are connected via a positive edge}, \\
0, & \text{otherwise}.
\end{cases}
\]

A path is a node sequence \( P = (v_1, v_2, \ldots, v_{l+1}) \), where \((v_i, v_{i+1}) \in E(G)\) for \( 1 \leq i \leq l \) and \( l \) is the length of path \( P \). If two users are not directly connected but are reachable from each other via a path, then we can obtain the potential relationship between them based on structural balance theory. The potential relationship may be friendship or an enemy relationship. The relationship matrix is denoted as \( R \), which only consists of entries of 1(friendly) and \(-1\)(hostile). For details, see figure 1. A path between \( u \) and \( v_3 \) is denoted as \((u, v_1, v_2, v_3)\), where \( l = 3 \) is the length of the path. According to structural balance theory, a balanced cycle in a signed network indicates that the product of all signs on the cycle is 1, namely, there are an even number of negative links on the cycle. The approach for determining the potential relationship between \( u \) and \( v_3 \) is based on structural balance theory. Hence, the sign of \( uv_3 \) should render balanced the cycle that is formed by combining \( uv_3 \) and the shortest published path. Thus, since the number of negative links on the path is even, then \( v_3 \) is a potential friend of \( u \). Similarly, \( v_4 \) is a potential enemy of \( u \). For simplicity, unless otherwise specified, we use friends (enemies) to refer to the directly connected and potential friends (enemies).
2.2. PID-mechanism

Suppose that information \( m \) originates from user \( u \). Here, we call user \( u \) the stakeholder of the message. It can be good (beneficial) or bad (detrimental) for \( u \). We denote the message as \( (g, b)_u \), in which \( g + b = 1 \), where \( g \) and \( b \in [0, 1] \) and \( g \) indicates the extent to which the message is good for user \( u \). Analogously, \( b \) indicates the extent to which the message is bad for user \( u \). If \( m \) is a relatively good message and \( g > 0.5 \) if the message is bad for user \( u \). In the special case of \( g = b = 0.5 \), the message is equivalent for every user in the social system.

The key assumption in this study is that if message \( m \) is good for user \( u \), then \( m \) is also good for its friends but bad for its enemies. For example, in figure 1, \( v_1 \) is \( u \)'s friend, and \( v_2 \) is an enemy of \( u \). The message \( m \) can be described as \( (g, b)_{v_1} \) and \( (b, g)_{v_2} \). The friends and enemies of user \( u \) differ in terms of attitude regarding message \( m \). The friends tend to publish a good message with a higher probability than the enemies. We define this personalized information diffusion mechanism, which is denoted as the PID-mechanism and belongs to a content-based information process.

As illustrated in figure 2, in the PID-mechanism, there are four states: unknown (U), insider (I), publisher (P) and forgetter (F). When user \( u \) publishes a message \( m \), the states of all of her/his neighbors are informed and change to insiders from unknowns. An insider will choose to publish the message with a probability of \( \xi \) according the relationship with the stakeholder of the message. In addition, the publisher will become a forgetter by removing the message at the next step.

In this study, the publishing probability \( \xi \) for user \( u \) is expressed as an increasing function of \( g \), \( \xi = f(g) \in [0, 1] \). In the first step, the stakeholder will publish the message with a probability of 1. We will discuss in detail the model in the following.

At \( t = 1 \), message \( m \) originates from user \( u \) and \( u \) publishes this message. At the next step, her/his informed friends have an intention of \( f(g) \) to publish this message, while the enemies are assumed to have an intention...
of $1 - f(g)$ to publish it. At the current step, the publisher can only inform her/his unknown neighbors of the message. At the next step, she/he will become a forgetter by removing the message. The insider will decide whether to publish the message at the current time. If so, her/his state will become published; otherwise, it will become a forgetter at the next step.

The main innovation of this study is that we assume the absolute property of the specified message $m$ does not change but the relative property might change according to the user’s relationship with user $u$. In signed social networks, the relationship between two individuals is determined by the directly connected links or the paths. During the diffusion process, the relative property of the message determines the publishing probability. In the final state of the diffusion process, only two classes of nodes are present in the network: the unknown and the forgetter.

### 2.3. Theoretical analysis

Each type of individual during the diffusion process can be divided into two classes: friends of $u$ and enemies of $u$. Here, we denote the numbers of the known, insiders, publishers and forgetters as $\mathcal{U}, \mathcal{I}, \mathcal{P}$ and $\mathcal{F}$, respectively. The friendly and antagonistic insiders (publishers) are denoted as $\mathcal{I}_f$ and $\mathcal{I}_e$ ($\mathcal{P}_f$ and $\mathcal{P}_e$), respectively. Hence, at step $t$, for example, the increment of the total number of insiders is $\Delta \mathcal{I}(t) = \mathcal{I}_f(t) + \mathcal{I}_e(t)$, and the increment of the total number of publishers is $\Delta \mathcal{P}(t) = \mathcal{P}_f(t) + \mathcal{P}_e(t)$.

We define the unsigned underlying structure of $G$ as $G^*$. The structure of $G^*$ is defined through its degree distribution $\{p_k, k = 0, 1, 2, \ldots\}$. In this section, we assume that the negative and positive edges are well mixed with probabilities of $p^-$ and $p^+$, respectively. Considering that $v$ is a friend of $u$, $v$’s neighbors are also $u$’s friends with a probability of $p^+$ and are $u$’s enemies with a probability of $p^-$. The case in which $v$ is $u$’s enemy is similar. Therefore, the relationship transmissibility matrix $T$ is

$$T = \begin{bmatrix} p^+ & p^- \\ p^- & p^+ \end{bmatrix}. \quad (1)$$

Let $\mathcal{I}(0) = \mathcal{I}_f(0) = 0$. At $t = 1$, $\mathcal{P}(1) = \mathcal{P}_f(1) = 1$ and the only publisher of the message is shakeholder $u$. At the same step, all the neighbors become insiders, and it follows that

$$\begin{bmatrix} \mathcal{I}_f(1) \\ \mathcal{I}_e(1) \end{bmatrix} = \sum_{k=0}^{+\infty} p_k k T \begin{bmatrix} \mathcal{P}_f(1) \\ \mathcal{P}_e(1) \end{bmatrix}, \quad (2)$$

where $k$ is the possibility degree of user $u$.

Since $u$’s friends and enemies differ in terms of intention to publish the message, we define the publishing matrix as

$$B = \begin{bmatrix} f(g) & 0 \\ 0 & 1 - f(g) \end{bmatrix}. \quad (3)$$

Hence, at step 2, the number of publishers satisfies

$$\begin{bmatrix} \mathcal{P}_f(2) \\ \mathcal{P}_e(2) \end{bmatrix} = \sum_{k=0}^{+\infty} p_k k B T \begin{bmatrix} \mathcal{P}_f(1) \\ \mathcal{P}_e(1) \end{bmatrix} = B \begin{bmatrix} \mathcal{I}_f(1) \\ \mathcal{I}_e(1) \end{bmatrix}, \quad (4)$$

For $t \geq 2$, we use a tree-like approximation to derive the increments of the insiders and publishers at each step. The degree of the neighbor of a randomly selected node follows the excess degree distribution, namely, $q_{uv} = k' p_v / \langle k' \rangle$, where $k'$ is the degree of the neighbor. The neighbor can only inform $k' - 1$ of its neighbors at the current step. Thus, the number of insiders at $t \geq 2$ is

$$\begin{bmatrix} \mathcal{I}_f(t) \\ \mathcal{I}_e(t) \end{bmatrix} = \sum_{k=0}^{+\infty} kp_v \langle k \rangle (k - 1) T \begin{bmatrix} \mathcal{P}_f(t - 1) \\ \mathcal{P}_e(t - 1) \end{bmatrix}, \quad (5)$$

Similarly, the number of publishers at step $t + 1$ is

$$\begin{bmatrix} \mathcal{P}_f(t + 1) \\ \mathcal{P}_e(t + 1) \end{bmatrix} = \sum_{k=0}^{+\infty} kp_v \langle k \rangle (k - 1) B T \begin{bmatrix} \mathcal{P}_f(t) \\ \mathcal{P}_e(t) \end{bmatrix} = B \begin{bmatrix} \mathcal{I}_f(t) \\ \mathcal{I}_e(t) \end{bmatrix}. \quad (6)$$
The total number of publishers up to step \( T \) is denoted as \( P^{\text{tot}}(T) \), which is calculated as

\[
P^{\text{tot}}(T) = \sum_{t=1}^{T} P(t) = \sum_{t=1}^{T} \left( P_{+}(t) + P_{-}(t) \right)
\]

\[
= \sum_{t=1}^{T} \left[ 1 \quad 1 \right] B \left[ I_{+}(t-1) \quad I_{-}(t-1) \right].
\]  

Similarly,

\[
I^{\text{tot}}(T) = \sum_{t=1}^{T} I(t) = \sum_{t=1}^{T} \left( I_{+}(t) + I_{-}(t) \right)
\]

\[
= \sum_{t=1}^{T} \left[ 1 \quad 1 \right] \left[ I_{+}(t) \quad I_{-}(t) \right].
\]

Compared with an insider, a publisher has a more significant influence on the diffusion of information. Here, we focus on the following question: what is the average probability for someone who knows the message to spread it? The proposition of insiders that publish the message until step \( T \), which is denoted by \( R(T) \), is calculated as

\[
R(T) = P^{\text{tot}}(T)/I^{\text{tot}}(T-1).
\]  

In the following, we provide an example to illustrate the relationship between the properties that are analyzed above.

### 3. Experiments on synthetic networks

There are many signed network models to generate synthetic signed networks [18, 22, 23]. For example, Wang et al treat the signed network as a special type of two-layer (i.e. positive and negative layers) network, and generate the two layers separately [18]. In the field of network science, two of the most popular network models are the Erdős–Rényi (ER) and scale-free (SF) network models. Here, we focus on the simplest signed networks. We use these two network models to generate synthetic unsigned networks, then randomly assign each link a positive sign with a probability of \( p^{+} \).

For Erdős–Rényi networks [24], set \( N = 10^4 \), and the average degree \( E[D] = 10 \). Let \( p^{+} = 0, 0.2, 0.5, 0.8 \) and 1, and vary \( g \) from 0 to 1 with an interval of 0.1. The degree distribution of the ER network can be well approximated by the Poisson distribution, especially when the network is sufficiently sparse. Therefore, the degree distribution of the unsigned underlying network is a Poisson distribution: \( p(k) = \lambda^k e^{-\lambda}/k! \), where \( \lambda = 10 \). For the publishing matrix \( B \), we set \( f(g) = g \) for simplicity. The theoretical results are obtained according to equations (6), (7) and (9) in the following discussion.

As described above, individuals who only obtain the information will not significantly impact the dissemination of information. However, the users who like to share the message will have an enormous influence on the diffusion scale. Therefore, the first question that arises regarding the proposed mechanism is as follows: with what probability will an insider publish the message? First, we investigate the evolution of \( R \) for various types of information. Considering \( p^{+} \in \{0, 0.2, 0.5, 0.8\} \) as examples, the top row of figure 3 shows how \( R \) evolves to the stable state as \( T \) increases, where the solid lines represent the simulation results and the dashed lines represent the theoretical results. \( R \) as a function of the diffusion step \( T \) can be well approximated by our theoretical result in equation (9) when \( T \) is relatively small. Since the theoretical analysis is based on the branching process, when \( T \) is large, the underlying structure is still a tree and the diffusion range is not constrained by the network size. However, the size of the synthetic network is constant, and as \( T \) increases, the global structure around user \( u \) will be considered, which is not tree-like. Hence, the theoretical analysis does not perform well when \( T \) is large. In addition, the theoretical result of \( R \) will converge to the same value for the cases of \( g = \alpha \) and \( g = 1 - \alpha \), where \( \alpha \in [0, 1] \). However, the process differs substantially between the cases of \( p^{+} > 0.5 \) and those of \( p^{+} < 0.5 \) in both the theoretical and simulation results. For \( p^{+} = 0.8 \), when \( g < 0.5 \), the value of \( R \) drops sharply to a value that is lower than the stable value and subsequently increases to the stable state. For \( g > 0.5 \), the value of \( R \) drops moderately to the stable state. The cases in which \( g < 0.5 \) require more steps to reach the stable state than those in which \( g > 0.5 \). In figure 4, except for the exceptional cases of \( p^{+} = 1 \) and \( p^{+} = 0 \), \( R \) is symmetric with \( g = 0.5 \) as the center when \( T \) is sufficiently large for \( R \) to become stable. The well-mixed edge signs can explain the symmetry of \( R \) with respect to the attribute of the message. Since the stable state of \( R \) does not increase with \( g \) when \( p^{+} > 0.5 \) and does not decrease with \( g \) when \( p^{+} < 0.5 \), even though an individual might have more directly connected friends than enemies, an insider might have a higher probability of publishing a relative bad message than a good message on average.
In previous studies on spreading models, the spreading capacity, or the number of individuals who become infected causes the most concern. Similarly, we concentrate on the size of the publisher group during the diffusion process. Since we study the tree-like network using branching theory, we only count the total number of publishers in the first 8 steps. Most diffusion processes stop after less than eight steps in synthetic networks. In addition, in the view of the phrase ‘six-degrees of separation’, eight steps are sufficient for a user to spread the information. As shown in the middle row of figure 3, the simulation results are consistent with the theoretical results when $T$ is small. The main reason the theoretical analysis does not perform well is that the branching process is not constrained by any constant, whereas the number of publishers in the simulation is restricted by the network size. For the cases with $p^+ < 0.5$, an extremely good or bad message is not easy to spread, and neutral information is easier to spread over a larger range. Similar to the observation on $R$, the number of publishers is also a nonmonotonic function of $g$. In terms of the absolute number of publishers, bad news may spread further than good news even if a user has more directly connected friends than enemies.

The last property that we investigate is the percentage of friends in the publisher group, which can be calculated according to equation (6). When diffusion step $T$ is sufficiently large, the friend proportion will converge to a stable value. As shown in figure 4(b), the observation is consistent with the intuition that friends will promote the diffusion of good messages while enemies will try to spread bad messages. However, the shape of the friend percentage as a function of message attribute $g$ differs among networks with different positive link proportions. The larger the value of $p^+$, the more sensitive the percentage is at the point $g = 0.5$. If $p^+ < 0.5$, then the proportion is more sensitive when $g$ is approximately 0 or 1. The bottom row of figure 3 shows the evolution of $P_f/P$. Similar to the discussion above, the theoretical analysis results accord with the simulation.
results when step $T$ is small. For the case of $p^+ = 0.5$, the simulation results are perfectly consistent with the theoretical results, even when $T$ is large.

Since degree distributions in real world networks mostly obey power law distributions, we conduct additional experiments on scale free networks. For SF networks [25,26], let $\lambda = 3$ and $\mathbb{E}[D] = 20$. Similarly, we set $p^+ = 0, 0.2, 0.5, 0.8$ and 1. The first observation is the rapid decline of $R$ at step 2 in left of figure 5 when $g < 0.5$, which indicates that most insiders will not publish a message that has already been transferred twice. This is mainly because when $p^+ < 0.5$, the majority of $u$’s 2nd-order neighbors are friends, and as stated above, friends will avoid spreading the user’s bad message. At $T = t$, the publishers are $(t - 1)$-order neighbors of user $u$. Therefore, when $T = 4$, $R$ starts increasing, which is because most insiders at the last step are $u$’s enemies and they prefer to spread $u$’s bad message.

In figure 6, we plot the percentages of friends and enemies with a distance of less than or equal to $d$. The explanation that is present above is further supported. Comparing the ER and SF networks, the increase in $R$ for SF starts from step 3, while for ER, the increment starts slightly later. From figure 6, the phase transition of the number of enemies or friends occurs at $d = 2$ in the SF networks, but at $d = 3$ in the ER networks.

A similar result for $p^+ > 0.5$ can be obtained. In addition, in signed networks, according to structural balance theory, when all links in the network are positive, all reachable individuals are $u$’s friends. However, if all links are negative, not all the other individuals are $u$’s enemies. There remain users with good relationships with $u$ since ‘the enemy of my enemy is my friend’.

4. Experiments on empirical networks

Finally, we select three real-world networks and explore how the intrinsic properties of real-world systems influence the characteristics of the proposed mechanism. The basic features of the three networks are listed in table 1. We compare the real-world networks with their randomized versions, which were constructed by randomly reshuffling the signs of links while maintaining the total degree of each node. Since the simulations are conducted on the largest connected component of each real-world network, the size, the number of positive links, and the number of negative links in the randomized networks are the same as those in the original.
Table 1. Basic features of three real-world signed networks

| Network           | # Nodes | # Links | # Positive links |
|-------------------|---------|---------|-----------------|
| Bitcoin alpha [27] | 3,783   | 13,867  | 12,724 (91.70%) |
| Bitcoin otc [28]  | 5,881   | 21,134  | 18,233 (86.27%) |
| Wikipedia [11]    | 7,118   | 100,355 | 78,440 (78.16%) |

Figure 7. Personalized information diffusion on real-world networks. The solid lines represent the results on original real-world networks and the dashed lines represent the results on their reshuffled networks.

Figure 8. Comparison of the average numbers of friends and enemies with a distance of $d$.

real-world networks. However, the positive and negative degree distributions differ; hence, the distribution of friends and enemies differ.

In the top row of figure 7, we show the final state of the proposed diffusion mechanism. The simulation results in the real-world networks are consistent with those in scale-free networks. For the ratio of the number of publishers to the number of insiders, the curves of $R$ as functions of $g$ cross each other at $g = 0.5$ and $R = 0.5$. Since $p^+ > 0.5$ for all networks, the larger $g$ is, the more likely insiders are to publish the message, which is consistent with previous observations.

According to figure 8, there are more friends in the original real-world networks than in the reshuffled networks within a distance of $d$. In contrast, there are more enemies in the reshuffled networks, especially for the two Bitcoin user trusting networks. Therefore, the message can be spread widely in the original real-world networks, as shown in figure 7(c). Analogously, there are more friends in the publisher group, (see figure 7(e)).

For the evolutions of $R$, $P_\text{tot}$ and $P_\text{tot}^+/P_\text{tot}$, we present two cases: $g = 0.2$ and 0.8. For the number of publishers $P_\text{tot}$ in figure 7(d), the phase transition occurs at $T = 4$ since the phase transition of the number of friends as a function of $d$ occurs at $d = 3$. When $g < 0.5$, the more (potential) enemies a user has, the more widely the message is spread. Similarly, the more (potential) friends a user has, the more widely a good message is spread. For the ratio of insiders who can publish the information and the percentage of friends in the publisher group, we obtain similar conclusions to those in synthetic networks.
5. Discussion

As we discussed above, the proposed PID-process belongs to a class of spreading models. A suitable spreading model is of academic significance in the study of vital node identification. First, the diffusion range of each node in the proposed model can be regarded as a benchmark metric for the evaluation of centrality metrics. In addition, we can use the proposed model to identify the critical nodes in signed networks. In this section, we mainly investigate the properties and characteristics of the top-ranked nodes regarding the diffusion range. The analysis is also conducted based on the three real-world networks that are described above.

Firstly, we identify the types of nodes that have a higher diffusion capacity in the proposed model. In figure 9, the $x$-axis corresponds to the distance, and the $y$-axis corresponds to the message property $g$. In the left panel of the figure, the color denotes the Kendall coefficient between the list of the number of (potential) friends with a distance if less than or equal to $d$ for each node and the diffusion capacity list for each node. For good information, the more friends a user has globally, the wider the information will be diffused. Analogously, in the right panel of the figure, we plot the Kendall coefficient between the number of enemies and the diffusion capacity. The more enemies a user has locally, the wider the bad message will be spread. In conclusion, the diffusion scale of good information depends on how many (potential) friends the stakeholder has across the network, while the diffusion capacity of bad information depends on the number of enemies the holder has in the nearby neighborhood. We obtain similar conclusions from the original networks and the sign-reshuffled networks; hence, the randomness of the sign will not substantially impact the properties of the proposed model from a global perspective.

From the above discussion, we do not observe any difference between the real-world networks and their reshuffled-versions. Next, we will explore the changes that occur during the reshuffling process. In figure 10, each point represents the numbers of publishers with each node as the information holder before the reshuffling and after the reshuffling, and the color represents different information property $g$. For most individuals, the total number of publishers remains unchanged before and after the reshuffling. However, for many nodes, the number of publishers changes substantially. The nodes on the $x$-axis indicate that after the reshuffling, the message cannot be spread. Similarly, the nodes on the $y$-axis indicate that the message cannot be spread widely, while after reshuffling, the information can be spread. Recall that the underlying network structure of a real-world network and those of reshuffled versions are the same. The reshuffling process only modifies the local signs of the directly connected links. This observation provides insights into information promotion and prohibition. By controlling the relationships with the directly linked neighbors, we can control the spread of various types of information, such as the rumors and epidemics. Suppose someone has a friend who has many enemies. These enemies will ‘help’ proliferate the bad information about him. Analogously, if someone has an enemy who has many opponents, then these opponents will help to promote good
6. Conclusions

In this article, we explored the spreading processes of personalized information in heterogeneous networks. We proposed a customized information diffusion model on signed social networks. The nodes in a signed network can be connected with two types of edges, namely, positive and negative, which correspond to friendly and rivalry relationships, respectively.

One of the most important theories that are related to signed networks is structural balance theory, which states that triangles that satisfy ‘the friends of my friends, along with the enemies of my enemies, are my friends’ are balanced; otherwise, they are unbalanced. For longer cycles, cycles with an even (odd) number of negative links are balanced (unbalanced). In terms of sociology, the balanced configurations are relatively stable, and the unbalanced configurations are more likely to break up. For a connected signed network, the relationship between any pair of nodes can be friendly or hostile. If the two nodes are not directly connected, then the relationship is determined by structural balance theory, namely, the sign of the imagined link should render the cycle balanced.

In the proposed model, the nodes have four states: unknown, informed, published, and forgotten. If a node is informed, we call it an insider. If a node published the message, we call it a publisher. We focus mainly on investigating the properties that are related to the insiders and the publishers. For a message $m$ with holder $u$, the publishing probability of an insider depends on the relationship with $u$. We theoretically and experimentally explored three properties of the proposed model: the probability of an insider publishing the message, the number of publishers as a function of step $T$, and the ratio of friends in the publisher group.

Theoretically, we found that even though an insider might have more directly connected friends than enemies, it might have a higher probability of publishing a relatively bad message than a good message on average. In terms of the absolute number of publishers, bad news may spread even further than good news. In addition, the larger $p^+$ is in the network, the more sensitive the friend percentage is around the point $g = 0.5$. If $p^+ < 0.5$, the percentage of friends in the publisher group is more sensitive around $g = 0$ and $g = 1$. For the simulations on synthetic networks, the results are well approximated by the theoretical results that are derived with branching theory when $T$ is small. In the experiments on real-world networks, we found that the nodes that have more (potential) friends across the network can spread good information more widely. However, individuals who have more enemies locally can spread bad information more widely.

Many questions arise that merit further investigation. In the experiments, we set the publishing probability as a function of $g$. A reasonable extension of the probability can be conducted, for example, $p = f(g,d)$, where $d$ is the distance between the user and the message holder. Many optimization problems that are worth exploring include how to add nodes or links to an existing signed network to maximize the diffusion range. In addition, we believe our proposed model can be applied in the task of network representation as an approach for sampling the neighborhood.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: http://snap.stanford.edu/data/index.html.
Acknowledgments

The authors were supported in part by National Natural Science Foundation of China under Grant Nos. 11631014, 11871311 and 12001324, in part by China Postdoctoral Science Foundation under Grant Nos. 2019TQ0188 and 2019M662315, in part by Shandong University multidisciplinary research and innovation team of young scholars under Grand No. 2020QNQT017.

ORCID iDs

Cunquan Qu https://orcid.org/0000-0001-8090-2538

References

[1] Zhang Z-K, Liu C, Zhan X-X, Lu X, Zhang C-X and Zhang Y-C 2016 Dynamics of information diffusion and its applications on complex networks Phys. Rep. 651 1–34
[2] Moreno Y, Nekovee M and Pacheco A F 2004 Dynamics of rumor spreading in complex networks Phys. Rev. E 69 066130
[3] Pastor-Satorras R and Vespignani A 2001 Epidemic spreading in scale-free networks Phys. Rev. Lett. 86 3200
[4] Moreno Y, Pastor-Satorras R and Vespignani A 2002 Epidemic outbreaks in complex heterogeneous networks Eur. Phys. J. B 26 521–9
[5] Szell M, Lambiotte R and Thurner S 2010 Multirelational organization of large-scale social networks in an online world Proc. Natl Acad. Sci. 107 13636–41
[6] Tang J, Chang Y, Aggarwal C and Liu H 2016 A survey of signed network mining in social media ACM Comput. Surv. 49 42
[7] Facchetti G, Iacono G and Altafini C 2011 Computing global structural balance in large-scale signed social networks Proc. Natl Acad. Sci. 108 20953–8
[8] Liu C, Liu J and Jiang Z 2014 A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks IEEE Trans. Cybern. 44 2274–87
[9] Heider F 1946 Attitudes and cognitive organization J. Psychol. 21 107–12
[10] Cartwright D and Harary F 1956 Structural balance: a generalization of Heider’s theory Psychol. Rev. 63 277
[11] Leskovec J, Huttenlocher D and Kleinberg J 2010 Signed networks in social media Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (New York: ACM) pp 1361–70
[12] Srinivasan A 2011 Local balancing influences global structure in social networks Proc. Natl Acad. Sci. USA 108 1751–2
[13] Kirkley A, Cantwell G T and Newman M 2019 Balance in signed networks Phys. Rev. E 99 012320
[14] Abell P and Ludwig M 2009 Structural balance: a dynamic perspective J. Math. Sociol. 33 129–55
[15] Marvel S A, Kleinberg J, Kleinberg R D and Strogatz S H 2011 Continuous-time model of structural balance Proc. Natl Acad. Sci. USA 108 1771–6
[16] Shi G, Altafini C and Baras J S 2019 Dynamics over signed networks SIAM Rev. 61 229–57
[17] Shi G, Prothier A, Johansson M, Baras J S and Johansson K H 2016 The evolution of beliefs over signed social networks Oper. Res. 64 585–604
[18] Wang H, Qu C, Jiao C and Ruszei W 2019 Self-avoiding pruning random walk on signed network New J. Phys. 21 035001
[19] Antia R, Regoes R R, Koella J C and Bergstrom C T 2003 The role of evolution in the emergence of infectious diseases Nature 426 658–61
[20] Eletreby R, Zhuang Y, Carley K M, Yajian O and Poor H V 2020 The effects of evolutionary adaptations on spreading processes in complex networks Proc. Natl Acad. Sci. USA 117 5664–70
[21] Liu C, Ma Y, Zhao J, Nussinov R, Zhang Y-C, Cheng F and Zhang Z-K 2020 Computational network biology: data, models, and applications Phys. Rep. 846 1–66
[22] Xiang B-B, Ma C, Chen H-S and Zhang H-F 2018 Reconstructing signed networks via Ising dynamics Chaos 28 123117
[23] Qu C and Wang H 2019 Impact of structural balance on self-avoiding pruning walk Physica A 524 362–74
[24] Erdős P and Rényi A 1960 On the evolution of random graphs Publ. Math. Inst. Hung. Acad. Sci. 5 17–60
[25] Barabási A-L and Albert R 1999 Emergence of scaling in random networks Science 286 509–12
[26] Caldarelli G, Capocci A, De Los Rios P and Munoz M A 2002 Scale-free networks from varying vertex intrinsic fitness Phys. Rev. Lett. 89 258702
[27] Kumar S, Spezzano F, Subrahmanian V and Faloutsos C 2016 Edge weight prediction in weighted signed networks IEEE 16th Int. Conf. on Data Mining (ICDM) (Piscataway, NJ: IEEE) pp 221–30
[28] Kumar S, Hooi B, Makhiia D, Kumar M, Faloutsos C and Subrahmanian V 2018 Rev2: fraudulent user prediction in rating platforms Proc. of the 11th ACM Int. Conf. on Web Search and Data Mining (New York: ACM) pp 333–41