mathematics

MDPI

Article

A Coupled Mathematical Model of the Dissemination Route of Short-Term Fund-Raising Fraud

Shan Yang, Kaijun Su, Bing Wang * and Zitong Xu

School of Resources and Safety Engineering, Central South University, Changsha 410083, China; yangshan@csu.edu.cn (S.Y.); 205512132@csu.edu.cn (K.S.); xuzitong@csu.edu.cn (Z.X.)
* Correspondence: safeboy@csu.edu.cn

Abstract: To effectively protect citizens' property from the infringement of fund-raising fraud, it is necessary to investigate the dissemination, identification, and causation of fund-raising fraud. In this study, the Susceptible Infected Recovered (SIR) model, Back-Propagation (BP) neural network, Fault tree, and Bayesian network were used to analyze the dissemination, identification, and causation of fund-raising fraud. Firstly, relevant data about fund-raising fraud were collected from residents in the same area via a questionnaire survey. Secondly, the SIR model was used to simulate the dissemination of victims, susceptibles, alerts, and fraud amount; the BP neural network was used to identify the data of financial fraud and change the accuracy of the number analysis of neurons and hidden layers; the fault-tree model and the Bayesian network model were employed to analyze the causation and importance of basic events. Finally, the security measures of fund-raising fraud were simulated by changing the dissemination parameters. The results show that (1) for the spread of the scam, the scale of the victims expands sharply with the increase of the fraud cycle, and the victims of the final fraud cycle account for 12.5% of people in the region; (2) for the source of infection of the scam, the initial recognition rate of fraud by the BP neural network varies from 90.9% to 93.9%; (3) for the victims of the scam, reducing fraud publicity, improving risk awareness, and strengthening fraud supervision can effectively reduce the probability of fraud; and (4) reducing the fraud rate can reduce the number of victims and delay the outbreak time. Improving the alert rate can reduce victims on a large scale. Strengthening supervision can restrict the scale of victims and prolong the duration of fraud.

Keywords: the SIR model; BP neural network; fault tree; Bayesian network; financial fraud

MSC: 91D99

check for updates

Citation: Yang, S.; Su, K.; Wang, B.; Xu, Z. A Coupled Mathematical Model of the Dissemination Route of Short-Term Fund-Raising Fraud. *Mathematics* **2022**, *10*, 1709. https://doi.org/10.3390/math10101709

Academic Editors: Gregorio Giménez Esteban and Javier Valbuena

Received: 5 April 2022
Accepted: 11 May 2022
Published: 17 May 2022

1. Introduction

In recent decades, crimes in the financial field have increased rapidly, among which fund-raising fraud has become one of the most harmful economic crimes. The research on the prevention methods of fund-raising fraud usually starts from three levels: (1) analysis of the measures against dissemination sources; (2) analysis of the measures against the dissemination process; and (3) analysis of the measures against fraud victims. For the research on the measures against dissemination sources, Massimo Bartoletti et al. [1] designed a group of features related to the Ponzi scheme classification (active days of the contract, maximum daily transactions, number of users, the average investment amount of users, etc.) and evaluated these features using their F-score, AUC, and other measurement indicators. The results show that the random forest algorithm can successfully identify 31 Ponzi schemes among 32 Ponzi schemes and has the best detection effect. Through the analysis and extraction of keywords, Yu Wenqiang et al. [2] matched the keywords with the source code of the contract to be tested and judged the fraud type of the contract by the logic of the transaction record. By using the improved Easy Ensemble algorithm, Zhou Yucai [3] obtained better performance in address detection of the Bitcoin Ponzi scheme.

Zhang Yanmei et al. [4] proposed a Ponzi-scheme-contract detection method based on a deep neural network. This method has a precision rate of 99.6% and a recall rate of 96.3%, which are superior to the existing methods. For the analysis of the measures against the dissemination process, Erhan Bayraktar et al. [5] studied victims' property losses using the Susceptible Infected Recovered (SIR) model from the perspective of behavioral dependence on propagation speed, herd immunity, and external blockades. According to the operation mechanism of network multi-level marketing (MLM), Liu Chao et al. [6] put forward the law of state transition and divided the network nodes into organization nodes, susceptible nodes, infected nodes, and removed nodes. Moreover, the SIR model of network MLMs was established, and the revenue evolution rules of MLM organizers and participants were obtained. For the analysis of the measures against fraud victims, Klafft [7] believed that in the process of the continuous integration of the internet and the financial industry, the problem of information asymmetry will continue to deepen and pose greater capital risks to investors. Marie Vasek et al. [8] analyzed the survival cycle data of the Ponzi scheme with the Cox proportional risk model. It was found that the more dissemination between swindlers and victims, the longer the survival cycle of the scheme. Paul et al. [9] held that a unified supervision institution should be established, or the existing Consumer Financial Protection Bureau (CFPB) should take on the responsibility of supervision. Fan Jianxing [10] claimed that the conditions of financial crimes can be controlled to prevent financial crimes, which is an important aspect of comprehensive treatment. Taking Baoding citizens as the research object, Fan Xing [11] carried out a multi-dimensional statistical analysis of the situation of financial fraud encountered by Bao-ding citizens via paper and online questionnaires.

In the first analysis method, a large number of identification features, evaluation indicators, and advanced machine learning methods are used to improve the accuracy of fraud identification. The second analysis method is rarely used, and research on the dissemination model of fund-raising fraud is very rare. The third analysis method has been widely used to analyze fraud problems from multiple angles, while incomplete analysis may be caused by this method.

Based on the existing research, fund-raising fraud was analyzed from the above three kinds of analysis methods in this study. Firstly, the SIR model was used to simulate fraud dissemination; then, the number of victims in different fraud cycles was obtained, and the impact of dissemination parameters on the dissemination process was studied. Secondly, the fault tree and the Bayesian network were combined to systematically analyze the reasons why victims were cheated and the importance of each basic event, and the BP neural network was employed to identify investment intentions by setting identification parameters. Finally, through these three methods, a comprehensive analysis of short-term fund-raising fraud was conducted in this study. Therefore, this study contributes to the analysis of fraud propagation, causes, and recognition.

2. Materials and Methods

2.1. Population Dynamics

2.1.1. SIR Model

The traditional SIR model is mainly applied to the transmission analysis of infectious diseases, while Liu Chao et al. [6] applied this model to the transmission analysis of network MLM transmission. In the SIR model [12–15], the population is divided into susceptibles, victims, and alerts. The number of susceptibles is marked as $S(t)$, representing the number of people who are not cheated but are likely to be cheated at time t. The number of victims is recorded as $I(t)$, indicating the number of people who have been cheated and will spread the fraud at time t. The number of alerts is recorded as $R(t)$, denoting the number of alerted people after removing the victims from the total population at time t. If the total population is $\boldsymbol{N(t)}$, then $\boldsymbol{N(t) = S(t) + I(t) + R(t)}$. The relevant assumptions are as follows:

1. Regardless of dynamic factors such as birth, death, and mobility, the population always maintains a constant, that is, $\boldsymbol{N(t) \equiv C}$.

2. Once a victim comes into contact with the susceptibles, the fraud must have a certain infectivity. It is assumed that in unit time $\boldsymbol{t}$, the number of susceptibles that a victim can infect is proportional to the total number of susceptibles $\boldsymbol{S(t)}$ in this environment, and the proportional coefficient is $\boldsymbol{\beta}$. Thus, the number of people infected by all victims in unit time t is $\boldsymbol{\beta S(t) I(t)}$.
3. At time $\boldsymbol{t}$, the number of alerts removed from victims in unit time is directly proportional to the number of victims, and the proportion coefficient is γ. The number of alerts removed per unit time is $\gamma \boldsymbol{I(t)}$.

Figure 1 shows the detailed dissemination mechanism of the SIR model.

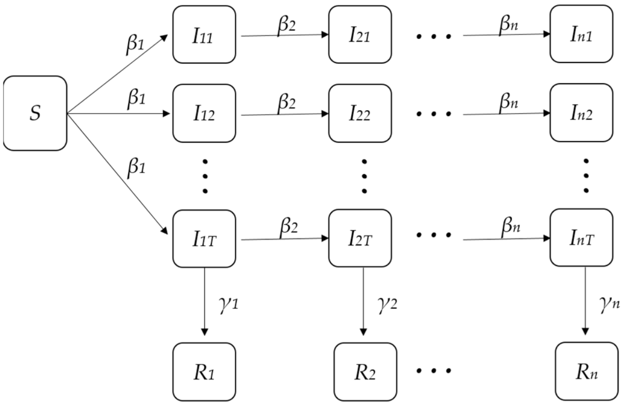


Figure 1. Propagation mechanism of the SIR model.

In Figure 1, $\boldsymbol{S}$ represents the susceptibles, $\boldsymbol{I_{ij}}$ represents the number of victims cheated on day j in the i-th fraud cycle, $\boldsymbol{R_i}$ represents the number of alerts in the i-th fraud cycle, $\boldsymbol{\beta_i}$ represents the fraud rate in the i-th fraud cycle, γ_i represents the alert rate in the i-th fraud cycle, and $\boldsymbol{T}$ represents the number of fraud cycles [16]. Through the above analysis, the differential equation of the i-th fraud cycle can be obtained in Equations (1)–(3):

$$\frac{dS(t)}{dt} = -\beta_i I(t) \frac{S(t)}{N(t)} \tag{1}$$

$$\frac{dI(t)}{dt} = \beta_i I(t) \frac{S(t)}{N(t)} - \gamma_i I(t) \tag{2}$$

$$\frac{dR(t)}{dt} = \gamma_i I(t) \tag{3}$$

The fraud cycles and total population of the model are set as follows: the complete fraud is composed of three fraud cycles; each fraud cycle lasts for 30 days, and a total of 90 days are involved in the fraud; the total population $\boldsymbol{N}$ involved in the fraud is 500,000.

2.1.2. Capital Flow and Profit of the Fraud

In the first fraud cycle [17], the victims cannot obtain any actual benefits; while in the second fraud cycle, the monthly investment profit is returned to the victims to promote the development scale of victims in the first and second fraud cycles and restrain the vigilance of the victims in the first fraud cycle. The purpose of the third fraud cycle is the same as that of the second fraud cycle, but the price of the new investment products promoted is higher. Based on the dissemination model, the capital flow of the fraud is analyzed, as shown in Figure 2.

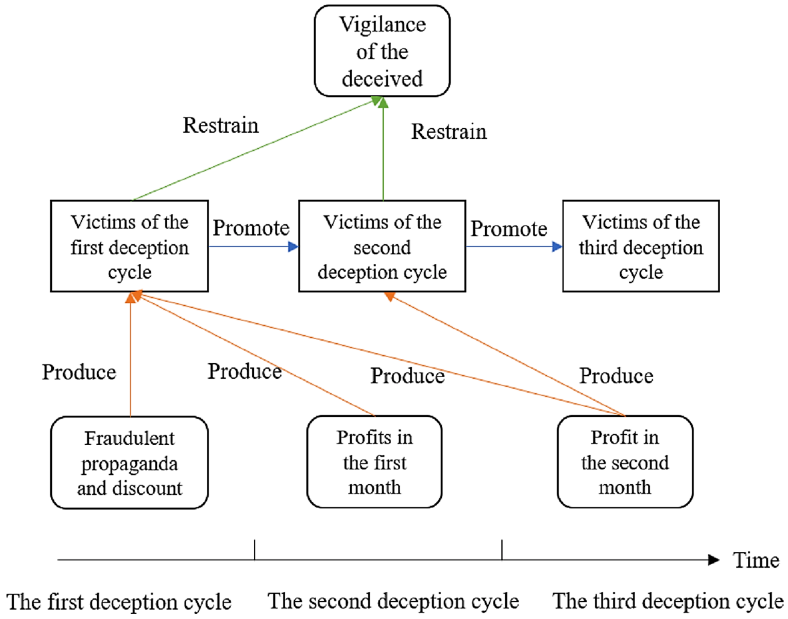

Vigilance of the deceived
Restrain
Restrain
Victims of the first deception cycle
Promote
Victims of the second deception cycle
Promote
Victims of the third deception cycle
Produce
Produce
Produce
Produce
Fraudulent propaganda and discount
Profits in the first month
Profit in the second month
Time
The first deception cycle
The second deception cycle
The third deception cycle

Figure 2. The capital flow of the fraud.

To effectively reflect the change of fraud profit [18] with the number of victims and alerts, the fraud profit in the n-th fraud cycle is divided into net fraud profit ($\boldsymbol{P_N}$), fraud profit loss ($\boldsymbol{P_L}$) [19], and maximum fraud profit ($\boldsymbol{P_M}$). The calculation equations of $\boldsymbol{P_N}$, $\boldsymbol{P_L}$, and $\boldsymbol{P_M}$ are as follows:

$$\boldsymbol{P_N = \sum_{i=1}^{n} C_i \int_{(n-i)T}^{(n+1-i)T} I^{'}(t)dt - A\sum_{i=1}^{n-1} C_i \int_{(n-i-1)T}^{(n-i)T} I^{'}(t)dt} \tag{4}$$

$$\boldsymbol{P_L = C_1 \int_{(n-1)T}^{nT} R^{'}(t)dt} \tag{5}$$

$$\boldsymbol{P_M = (P_N)_{max}} + (\boldsymbol{P_L})_{\boldsymbol{max}}, \tag{6}$$

where $\boldsymbol{C_i}$ is the average investment amount of the $\boldsymbol{i}$-th product, $\boldsymbol{A}$ is the monthly interest rate of the investment, $\boldsymbol{I'(t)}$ is the derivative of the number of victims to time, and $\boldsymbol{R'(t)}$ is the derivative of the number of alerts to time.

The net fraud profit $\boldsymbol{P_N}$ consists of two parts. The first part is the investment amount provided by the victims in different projects in $\boldsymbol{n}$ cycles. The second part is the monthly investment profit in $\boldsymbol{n} - 1$ cycles spent to maintain the fraud. The net fraud profit in this fraud cycle is calculated by subtracting the two parts. The fraud profit loss $\boldsymbol{P_L}$ comes from the amount of investment taken away by the victims who leave the fraud. The maximum fraud profit $\boldsymbol{P_M}$ comes from the sum of the maximum loss of the fraud profit and the maximum net fraud profit. To facilitate the calculation, it is stipulated that only the victims who do not receive the monthly profit can get out of the fraud, and the victims who receive the monthly profit continue to invest in the next fraud cycle.

2.2. BP Neural Network

BP neural network [20,21] is a kind of a neural network system that uses error back-propagation to realize feedforward correction. It is composed of the input layer, hidden layer, and output layer. Each layer has a different number of neurons, and more accurate results can be obtained by increasing the number of hidden layers of neurons. To identify whether a short-term financial investment is a financial fraud, a BP neural network needs to be trained with the data of input parameters and output parameters. In this study, the investment amount and monthly interest rate are taken as the input variables, expressed as $\boldsymbol{X} = \{x_1,x_2\}$, and the class is taken as the output variable, expressed as $\boldsymbol{Y} = \{y\}$. Four layers of the neural network are set up, including 1 input layer, 2 hidden layers, and 1 output layer; the number of variable nodes is $\boldsymbol{L}$ = 2 in the input layer; the number of variable nodes is $\boldsymbol{M}$ = 1 in the output layer; and the number of nodes N in the hidden layer is calculated by Equation (7), where $\boldsymbol{\alpha}$ ranges from 0–120 [22]. Figure 3 shows the simulated neural network structure.

$$\boldsymbol{N}=\sqrt{\boldsymbol{L}+\boldsymbol{M}}+\boldsymbol{\alpha} \tag{7}$$

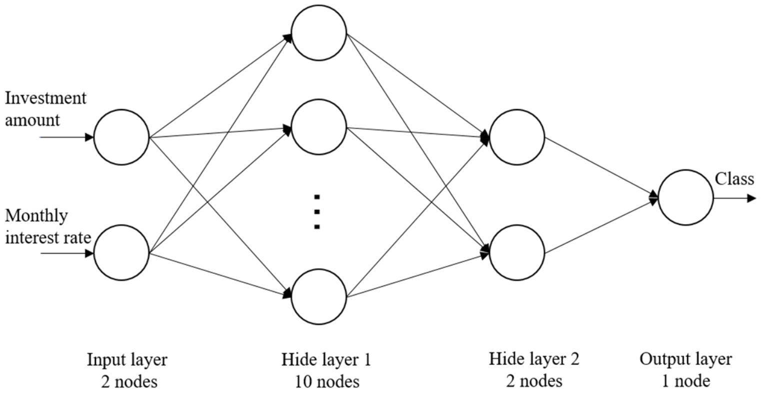


Figure 3. Structure diagram of the BP neural network.

The BP neural network model is used for data identification. The functional relationship [20] between class $\boldsymbol{Y}$ and input variable $\boldsymbol{X}$ can be expressed as:

$$\boldsymbol{Y}=\boldsymbol{F}(\boldsymbol{X})=\boldsymbol{\phi}\ [b_k+\textstyle\sum_{n-1}^{N} W_n\varphi(b_n+\sum_{l=1}^{L} w_{ln}x_l)], \tag{8}$$

where w_{ln} is the connection weight between the input layer and the hidden layer, $\boldsymbol{W_n}$ is the connection weight between the hidden layer and the output layer, $\boldsymbol{b_n}$ is the threshold of the hidden layer, $\boldsymbol{b_k}$ is the threshold of the output layer, and $\boldsymbol{\varphi(\cdot)}$ is the transfer function of the hidden layer. The activation function of the hidden layer is the hyperbolic tangent S-shaped function, the activation function of the output layer is linear, and the hidden layer is composed of hidden layer 1 (10 nodes) and hidden layer 2 (2 nodes). The gradient descent adaptive learning rate is adopted for the training function, and the maximum number of iterations is set as 600.

2.3. The Fault Tree and Bayesian Network

2.3.1. Basic Principles of the Fault Tree

Fault tree analysis (FTA) is a logical directed graph from result to cause. It can analyze the causes of accidents [23–25]. It is assumed that $\boldsymbol{X_i}$ and T represent the variables of basic events and top events, and 0 or 1 can be taken to represent the occurrence and non-occurrence of events, respectively.

$$\boldsymbol{P_{and}}=\prod_{i=1}^{n}\boldsymbol{P}(\boldsymbol{X_i}=1) \tag{9}$$

$$\boldsymbol{P_{or}}=1-\prod_{i=1}^{n}\boldsymbol{P}(\boldsymbol{X_i}=1) \tag{10}$$

The top event T will be determined by the state of basic events in the fault tree. When the top event T occurs (T = 1), the corresponding state of the basic event is $\boldsymbol{X_i}$ = 1. When $\boldsymbol{X_i}$ only takes 0 and 1, the probability calculation formulas of and-gate and or-gate are obtained, as shown in Equations (9) and (10). Moreover, $\boldsymbol{P(X_i = 1)}$ represents the probability of the occurrence of the basic event $\boldsymbol{X_i}$.

2.3.2. Basic Principles of Bayesian Network

Bayesian network [26–29] is a representation and reasoning model of uncertain knowledge based on graph theory and probability theory. It can be expressed as N = <(X, T), P>, where $\boldsymbol{X}$=$\{x_1,x_2\ldots x_n\}$ is the combination of nodes, and T is the set of oriented edges and indicates that there is a causal relationship between related variables. This network is suitable for expressing and analyzing events with uncertain probabilities. The Bayesian network represents the relationship between parent nodes $\boldsymbol{\pi(x_i)}$ and child nodes $\boldsymbol{x_i}$. At the same time, child nodes and other unrelated nodes $\boldsymbol{A(x_i)}$ are independent, as shown in Equation (11).

$$\boldsymbol{P(X_{€} \mid \pi(€), A(X_i)){=}P(X_i \mid \pi(X_i))} \tag{11}$$

The structure of learning of a Bayesian network is to establish the relationship between nodes. In this study, the fault tree method is used to construct a Bayesian network. Through the cause analysis of fault events, the basic events and intermediate events leading to the risk occurrence are found to build the fault tree, and the Bayesian network is generated from the fault tree [30]. Due to the conditional independence between basic events, the joint probability formula can be shown as in Equation (12).

$$\boldsymbol{P(X){=}P(X_1, X_2\ldots X_n){=}\prod\nolimits_{i=1}^{n} P(X_i \mid \pi(X_i)).} \tag{12}$$

The events in the fault tree correspond to those in the Bayesian network one by one, and the causality of the events remains unchanged. In other words, the dependency among each root node, an intermediate node, and a leaf node in the fault event is used as the prior knowledge in the Bayesian network to obtain the conditional probability of all nodes.

2.3.3. Importance Calculation

Structural importance analysis [31] is used to analyze the importance of each basic event from the structure of the fault tree. Specifically, without considering the probability of occurrence of each basic event or assuming that the probability of occurrence of each basic event is the same, the order of the structural importance of each basic event is analyzed and arranged to understand the influence of each basic event on the occurrence of the top event.

In fault tree analysis, each basic event has two states: (1) the basic event can occur, i.e., $\boldsymbol{X_i}$ = 1; and (2) the basic event cannot occur, i.e., $\boldsymbol{X_i}$ = 0. The different combinations of the states of different basic events constitute the different states of the top event, i.e., $\boldsymbol{\Phi(X)}$ = 1 or $\boldsymbol{\Phi(X)}$ = 0. When the state of a basic event $\boldsymbol{X_i}$ changes from 0 to 1 (i.e., 0i → li) and the states of other basic events remain unchanged, the state of the top event may change in four cases [32]:①the top event remains unchanged in the state of 0 (i.e $\boldsymbol{\Phi(X)}$ = 0 → $\boldsymbol{\Phi(X)}$ = 0);②the top event changes from 0 to 1 (i.e $\Phi(X)$ = 0 → $\Phi(X)$ = 1);③the top event remains in the state of 1 (i.e., $\boldsymbol{\Phi(X)}$ = 0 → $\boldsymbol{\Phi(X)}$ = 1); and④the top event changes from 1 to 0 (i.e $\boldsymbol{\Phi(X)}$ = 1 → $\boldsymbol{\Phi(X)}$ = 0). In these four cases, ① and ③ do not change with the basic events, while the top event in ④ changes from occurrence to non-occurrence [33]. Only the top event in ② changes from non-occurrence to occurrence, indicating that the top event is affected by the basic event. If specific measures are taken to prevent the occurrence of the basic event, the top event will not occur, and the fault can be avoided. This is the significance of analyzing the importance of structure. The number of mutually incompatible combinations of $\boldsymbol{n}$ basic events is $\boldsymbol{2^n}$ in total. When the basic event $\boldsymbol{X_i}$ is taken as the change object and the state of the other $\boldsymbol{(n-1)}$ basic events remains unchanged, $\boldsymbol{2^{n-1}}$ events can be generated. In these $\boldsymbol{2^{n-1}}$ events, the ratio of the number of events in line with ① to the total event in the control group is the structural importance coefficient of the event $\boldsymbol{X_i}$ [34]. This coefficient can be expressed by Equation (13).

$$\boldsymbol{I_{\Phi}(i)}{=}\frac{1}{2^{N-1}}\sum[\boldsymbol{\Phi}(1_i,X)-\boldsymbol{\Phi}(0_i,X)] \tag{13}$$

The importance of the probability [34] of the occurrence of basic events is called the probability importance. Its function is to calculate the probability importance coefficient.

According to the result, reducing the occurrence probability of a basic event with a high probability can reduce the occurrence probability of the top event. The probability function of the top event occurrence $\boldsymbol{Q}$ is a multivariate linear function, and $\boldsymbol{q_i}$ is an independent variable. The probability importance coefficient $\boldsymbol{I}$ of the basic event [35] can be obtained by calculating the partial derivative of the multivariate function, as shown in Equation (14):

$$\boldsymbol{I_q(i)}=\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q_i}}, \tag{14}$$

which shows the ratio of the change rate of the occurrence probability [36] of the top event to the change rate of the occurrence probability of the basic event, i.e., the critical importance $\boldsymbol{C_i}$. Critical importance [37] can reflect the importance of each basic event from the perspective of sensitivity and probability, as shown in Equation (15).

$$\boldsymbol{I_c(i)}=\frac{\boldsymbol{q_i}}{\boldsymbol{Q}}\boldsymbol{I_q(i)} \tag{15}$$

2.4. Description of the Survey

A self-administered questionnaire was distributed to over 1000 adults in Changsha, China. The snowball sampling approach [38] was used. Researchers contacted college students in different disciplines and asked them to recruit subjects from among their acquaintances. Researchers also described the purpose of the study to the college students and explained how to complete the questionnaire. Finally, a total of 1200 questionnaires were collected, of which 1032 were complete, and the effective rate of the questionnaire was 86%. A total of 258 participants (25%) had questionnaire results that indicated a loss and that they had been victims of at least the third fraud cycle, most of them were male (16.1%) and their ages ranged from 31 to 50 (19.4%).

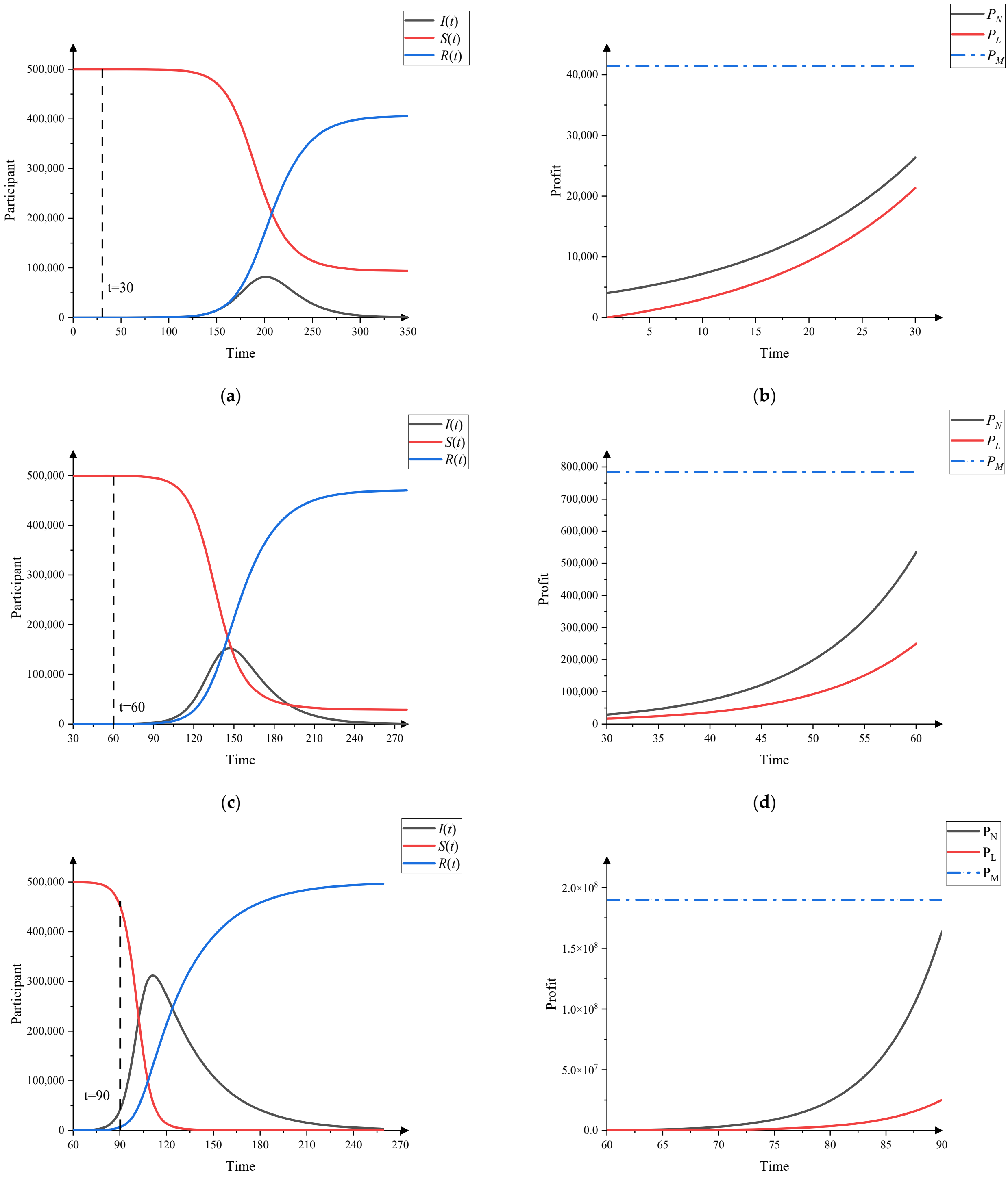
The items in the questionnaire survey included personal characteristics, investment management, and cognition. The specific questionnaire is attached in Table A1. The Chi-square test [39] was used to compare the victims who suffered losses with those who did not. It should be noted that all questionnaires were assessed by the self-perception of the participants. The descriptive statistics of corresponding variables are available in Table A2. In general, victims who suffered losses were more associated with receiving investment information, the amount of the first investment, and the withdrawal of investment within the third month than victims who did not suffer losses. The meanings and sources of the research data parameters used in this study are listed in Appendix C.

3. Results and Discussions

3.1. Dissemination Simulation of Fraud

For convenience, the parameters of the SIR model are listed in Table A3 of Appendix C. The numerical simulation of short-term fund-raising fraud is carried out, and the number of victims, susceptibles, and alerts in the same region in three fraud cycles and the changes in fraud profits are obtained. Figure 4 shows the simulation results.

In the first fraud cycle, the fraud rate is the lowest, and the alert rate is the highest; thus, the number of victims increases slowly with a smaller property loss. Figure 4a,b show the simulation results of the first fraud cycle. According to the dissemination simulation within 30 days, only 7 people fall into the fraud and 5 people break away from the fraud by the 30th day. From the whole dissemination process, the maximum number of victims reaches 80,000, which is less than the minimum number of susceptibles of 95,000. When the dissemination time reaches 200 days, the net amount of fraud in the first fraud cycle reaches 21,000 yuan, and the amount of fraud lost reaches 15,000 yuan. At this stage, the property loss caused by the fraud and the dissemination degree of the fraud is low, and the fraud is easy to be perceived. Moreover, the fraud cannot be disseminated on a large scale, and the dissemination of the fraud will decline in a very short time and die out by itself; it is difficult for the fraud to be detected by the regulatory authorities due to the small number of victims.



(**a**) (**b**)

(**c**) (**d**)

(**e**) (**f**)

Figure 4. Fraud spread and fraud profit. (**a**) Participants in the first cycle of fraud; (**b**) Profits from the first cycle of fraud; (**c**) Participants in the second cycle of fraud; (**d**) Profits from the second cycle of fraud; (**e**) Participants in the third cycle of fraud; (**f**) Profits from the third cycle of fraud.

In the second fraud cycle, the monthly investment profit is paid to develop the new victims. Therefore, the probability of being defrauded is higher than before, and the rate of breaking away from the fraud is lower than before. Figure 4c,d show the simulation results in the second fraud cycle. According to the dissemination simulation, within 60 days, only 141 people fall into the fraud, and 73 people break away from the fraud by the 60th day. From the whole dissemination process, the maximum number of victims reaches 150,000, and the arrival time of the peak fraud dissemination is 150 days. In the second fraud cycle, the net amount of fraud reaches 550,000 yuan, and the amount of fraud lost reaches 240,000 yuan. In this stage, a large scale of fraud and great losses can be caused. Compared with the first fraud cycle, the arrival time of peak fraud dissemination is 50 days earlier, indicating that fraud can erupt more quickly and form a rapid spread. At this time, the number of susceptibles, victims, and alerts is close, indicating that large-scale fraud dissemination has been developed, and relevant measures need to be taken to intervene, otherwise, large property losses can be caused. However, the minimum number of susceptibles is still about 30,000, which suggests that the dissemination of fraud does not cover the whole region.

In the third fraud cycle, the returned profit and publicity in the first two months lead to a further increase in the fraud rate; the residents' vigilance is weakened, which leads to a decrease in the alert rate. Figure 4e,f show the simulation results in the third fraud cycle. According to the dissemination simulation, within 90 days, the number of people who fall into the fraud is 41,048, and the number of people who break away from the fraud is 6228 by the 90th day. From the whole dissemination process, the number of maximum victims is 310,000; the arrival time of peak fraud dissemination is 115 days, the net amount of fraud in the third fraud cycle is 550,000 yuan, and the amount of fraud lost is 160 million yuan. The fraud at this stage begins to disseminate like influenza and causes huge losses. However, before the returned profit in the fourth month, victims rarely detect the fraud due to the investment profit. Compared with the second fraud cycle, the arrival time of the peak fraud dissemination is 35 days earlier in the third cycle. At this time, the number of susceptibles is close to 0, and the number of victims is more than twice that of alerts. It suggests that the residents of the whole region have participated in the fraud, and serious losses can be caused in this region. Moreover, strict measures must be taken to stop the dissemination of the fraud before and during its development.

3.2. BP Neural Network Analysis

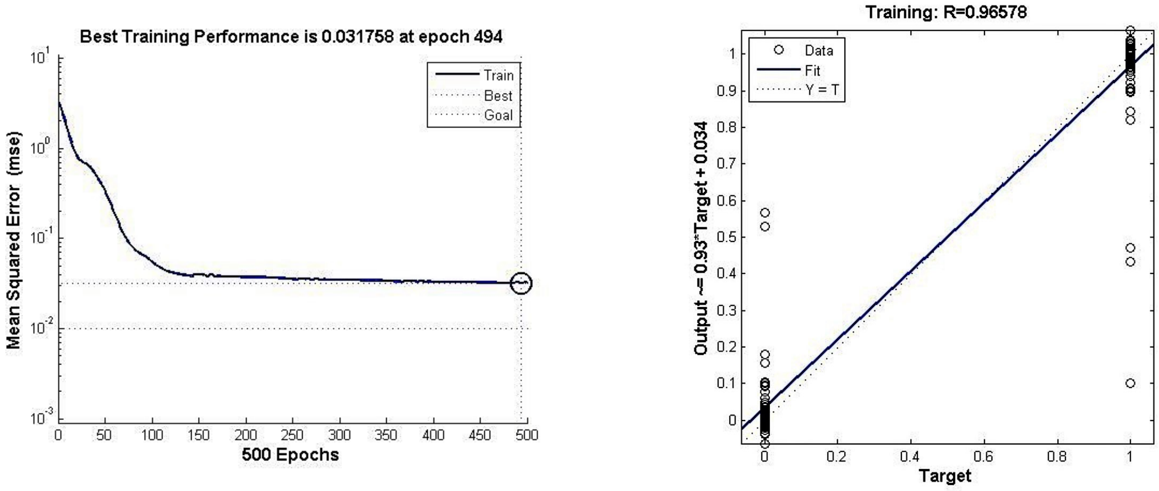
3.2.1. BP Neural Network Identification Analysis

In the BP neural network, 60% of the 1032 data records were collected in the survey as the training set (Table A4), and 40% as the test set (Table A5). The identification criteria in this study refer to relevant practical cases [40]. The input layer is the row matrix of the sample investment amount and monthly interest rate, the output layer is the row matrix of the sample class, and the output results are obtained, as shown in Table 1.

In the first fraud cycle, the fraud rate is the lowest, and the alert rate is the highest; thus, the number of victims increases slowly with a smaller property loss. Figure 4a,b show the simulation results of the first fraud cycle. According to the dissemination simulation within 30 days, only 7 people fall into the fraud and 5 people break away from the fraud by the 30th day. From the whole dissemination process, the maximum number of victims reaches 80,000, which is less than the minimum number of susceptibles of 95,000. When the dissemination time reaches 200 days, the net amount of fraud in the first fraud cycle reaches 21,000 yuan, and the amount of fraud lost reaches 15,000 yuan. At this stage, the property loss caused by the fraud and the dissemination degree of the fraud is low, and the fraud is easy to be perceived. Moreover, the fraud cannot be disseminated on a large scale, and the dissemination of the fraud will decline in a very short time and die out by itself; it is difficult for the fraud to be detected by the regulatory authorities due to the small number of victims.

Table 1. Identification results of test set data.

Number	Output		Class	Number	Output		Class
	2	1			2	1	
1	0.6811	0.3189	2	…	…	…	…
2	1.0002	−0.0002	2	399	1.0103	-0.0103	2
3	0.8709	0.1291	2	400	−0.0361	1.0361	1
4	0.0208	0.9792	1	401	1.0482	−0.0482	2
5	0.943	0.057	2	402	1.0132	−0.0132	2
6	0.9951	0.0049	2	403	0.5894	0.4106	2
7	1.0499	−0.0449	2	404	0.972	0.028	2
8	0.9951	−0.0012	2	405	1.0154	−0.0154	2
9	1.0012	−0.0336	2	406	1.003	−0.003	2
10	1.0336	−0.0336	2	407	0.9968	0.0032	2
11	−0.032	1.032	1	408	0.9363	0.0637	2
12	0.0051	0.9949	1	409	1.0121	−0.0121	2
13	0.9985	0.0015	2	410	−0.0081	1.0081	1
14	1.0038	−0.0038	2	411	0.9961	0.0039	2
15	0.0056	0.9944	1	412	−0.0137	1.0137	1
…	…	…	…	413	1.0056	−0.0056	2



(**a**) (**b**)

Figure 5. *MSE* of training and correlation coefficient diagram. (**a**) Image of the effect of epoch on mean square error; (**b**) Regression equation image of output data.

The correlation coefficient is R = 0.96578, indicating that the correlation between the output data and the target is very strong. The recognition accuracy of the test set data reaches 90.9%. It indicates that the fund-raising fraud can be preliminarily identified through the BP neural network, and a judgment can be formed as to whether the investment project is likely to be a fraud. In addition, the *T*-test and F test were used to verify the significance of the regression equation to ensure the reliability of the data relationships. Statistical test results are shown in Table 2.

Table 2. Correlation parameters of regression equation test.

		Regression Statistics			
Multiple R 0.979	R Square 0.959	Adjusted R Square 0.959	Standard error 0.09934		
		Analysis of variance			
	SS	df	MS	F	Significance F
Regression analysis	13.53	1	13.53	1370.903	0.00
Residual error	0.572	58	0.01		
Total	14.102	59			
		Coefficient			
	Coefficients	Standard error	t	*p*-value	
Intercept	0.034	0.018	1.355	0.181	
Output data	0.959	0.026	37.026	0.00	

According to the regression equation analysis of the data in Table 2, the F value of the equation is 1370.903, and the significance test value is Significance F is 0.00 < 0.001. The equation has passed the significance test, which, on the whole, indicates an obvious linear relationship between the dependent and independent variables. The R square and adjusted R square values are 0.979 and 0.959, respectively, indicating that the model has a high degree of fitting, and the regression equation is very representative. As for the relationship between the type of independent variable and the output value of a dependent variable, it can be seen from the regression analysis in Table 3 that the regression coefficient of the type of independent variable is 0.959, the T value is 37.026, and the *p*-value of T is 0.00 < 0.001, which passes the *T*-test.

Table 3. Event name and probability.

Event Number	**Event Name**	**Event Probability**	
		Bayesian Network	**Fault Tree**
M_1	Patsy investment	0.172	0.508
M_2	Cannot rational investment	0.0797	0.053
T	Financial fraud causes losses	0.0498	0.0027

3.2.2. The Influence of BP Neural Network Parameters

The influence of a parameter change on identification is studied by changing the parameters of the BP neural network [41]. The operation steps of 100, 200, 300, 400, 500, and 600 are used to obtain the images of the change in fraud recognition rate and error rate, as shown in Figure 6a.

The variation range of the recognition rate is small, which is relatively stable at around 90.9%. With the increase of the operation steps, the recognition rate changes between 90.9–93.9%, and the recognition rate decreases to 87.87% at operation steps of 600. The error rate decreases gradually with the increase of the operation step, indicating that, with the increase of the operation step, the recognition rate is still accurate. When the operation steps are 500 and 600, the error rate is stable at 4.5%. The recognition rate and error rate of fraud events are obtained by using 10, 20, 40, 80, and 160 hidden neurons, respectively, as shown in Figure 6b. With the increase in the number of hidden neurons, the recognition rate is relatively stable, maintained at 90.9%, and the error rate changes from 6% to 3%. Using neural networks to identify data can improve the identification rate to 90.9%, which is far greater than the fraud rate of 13.1%.

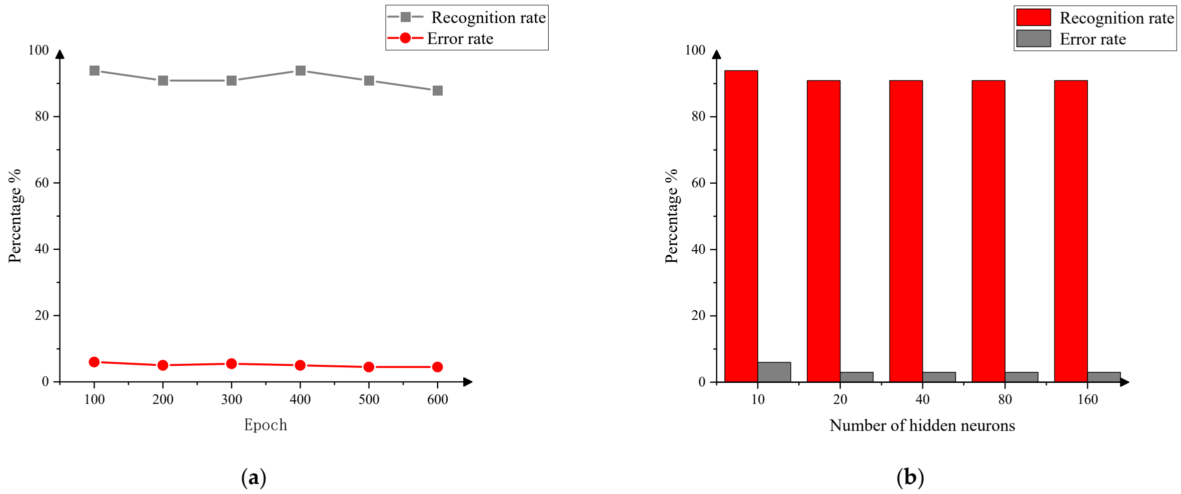


Figure 6. The effect of epoch and number of hidden neurons on fraud identification. (**a**) The effect of epoch on recognition rate; (**b**) The effect of number of hidden neurons on recognition rate.

3.3. Comprehensive Analysis of Fault Tree and Bayesian Network

3.3.1. Probability Calculation of Fault Tree and Bayesian Network

The dangerous and harmful factors of financial fraud are identified and analyzed through the compilation principle of the fault tree, and the causes of fraud losses are divided into direct causes (including human unsafe behavior, unsafe state of things, management errors) and indirect causes. Finally, it is concluded that the occurrence of fraud is caused by the following reasons: patsy investment, lack of risk awareness, and lack of fraud supervision. The probability of the basic events of the fault tree is shown in Table A6. A fault tree for fund-raising fraud is constructed by using the cause analysis of the statistical fraud events, as shown in Figure 7. The Bayesian network structure of fund-raising fraud can be obtained from the constructed fault tree. The conditional probability distributions of $\boldsymbol{M_2}$, $\boldsymbol{M_1}$, and $\boldsymbol{T}$ are shown in Tables A7–A9. Therefore, we can obtain the structure of the Bayesian network and the probability of event nodes, as shown in Figure 8.

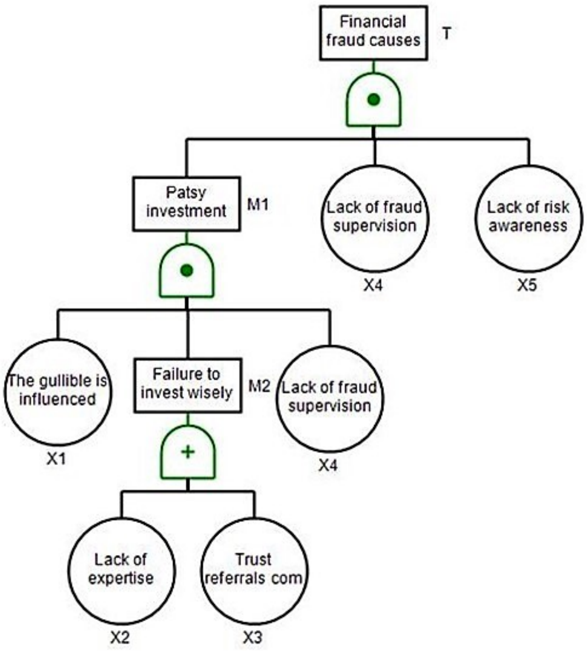


Figure 7. The fault tree of fund-raising fraud.

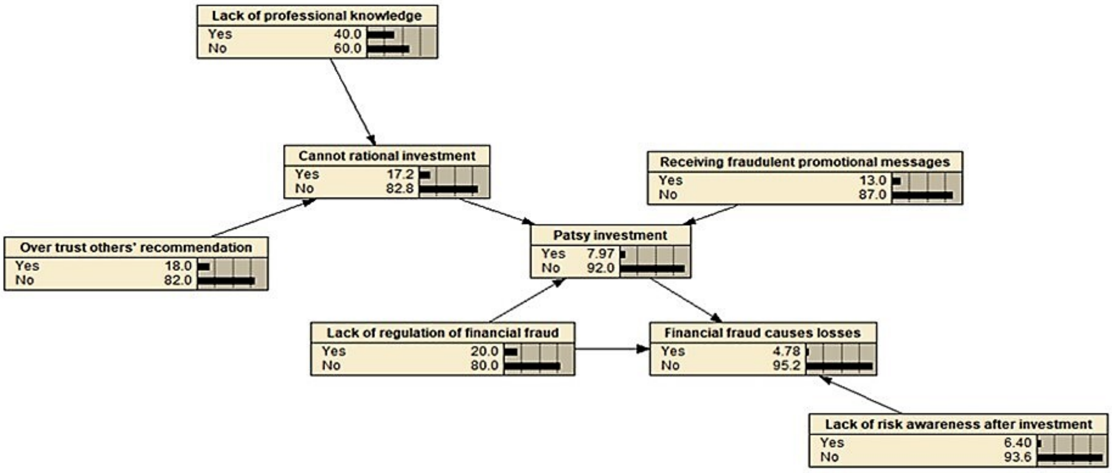

Lack of professional knowledge
Yes 40.0
No 60.0
Cannot rational investment
Yes 17.2
No 82.8
Receiving fraudulent promotional messages
Yes 13.0
No 87.0
Over trust others' recommendation
Yes 18.0
No 82.0
Patsy investment
Yes 7.97
No 92.0
Lack of regulation of financial fraud
Yes 20.0
No 80.0
Financial fraud causes losses
Yes 4.78
No 95.2
Lack of risk awareness after investment
Yes 6.40
No 93.6

Figure 8. The Bayesian network of fund-raising fraud.

The structural expression of the top event on the fault tree [23] is shown in Equation (16).

$$T{=}X_1{\cdot}X_2{\cdot}X_4{\cdot}X_5{+}X_1{\cdot}X_3{\cdot}X_4{\cdot}X_5 \tag{16}$$

Through the two different methods, the probability of an intermediate event and a top event can be calculated according to Equations (9)–(12), and the calculation results of the two methods are shown in Table 3.

There is little difference in the probability of the intermediate event M_2 calculated by the two methods. The probability of the intermediate event M_1 calculated by the fault tree is more than twice that when calculated by the Bayesian network, and the probability of the top event calculated by the Bayesian network is more than five times that calculated by the fault tree. Compared with the fault tree, the calculation of the Bayesian network does not focus on the occurrence of a single basic event [29] but on the posterior probability under the condition of a pre-event. Therefore, the calculation of the Bayesian network is more reliable. Although the probability of loss caused by fraud is only 0.0498, it will still cause great loss with the increase in transmission times and fraud amount. Therefore, it is necessary to analyze the importance of basic events to avoid the occurrence of losses.

3.3.2. Importance Analysis of Basic Events

To analyze the importance of the basic event, the structural importance, critical importance, and probability importance of the basic event are calculated, as shown in Equations (13)–(15). The structural importance, critical importance, and probability importance of the fault tree are shown in Figure 9.

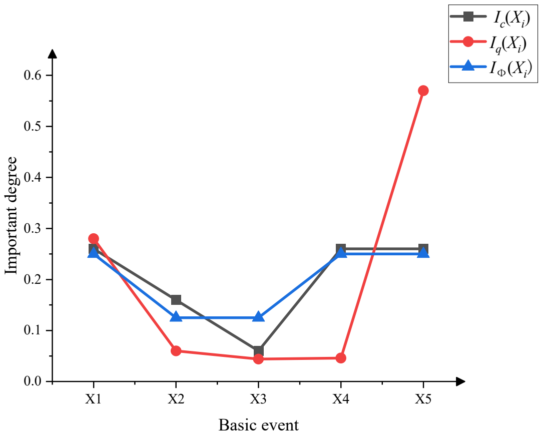


Figure 9. Comparison of three levels of importance.

According to the analysis of the structural importance, X_1, X_4, and X_5 have the same and largest structural importance. This indicates that, in terms of the structure of the fault tree, the occurrence of X_1, X_4, and X_5 has an important impact on the occurrence of the fault. Therefore, priority can be given to X_1, X_4, and X_5 in taking measures to avoid the fault occurrence. Through the analysis of the probability importance, it can be found that X_5 has the largest probability importance, followed by X_1. In other words, under the influence of the occurrence probability of basic events, X_5 is the most important to the fault occurrence, and measures can be taken to reduce the probability of X_5 to avoid the fault. According to the analysis of critical importance, X_1, X_4, and X_5 have greater structural importance. It suggests that, under the influence of the probability and sensitivity of basic events, the occurrence of X_1, X_4, and X_5 has an important impact on the fault occurrence. To avoid the occurrence of the fault, priority can be given to X_1, X_4, and X_5 when taking measures against the fault.

Based on the above analysis, the occurrence of basic events X_1, X_4, and X_5 has a significant impact on the occurrence of fund-raising fraud. Corresponding security measures should be taken to reduce the fraud rate, improve the alert rate, and strengthen the management of financial fraud, so as to reduce the occurrence possibility of X_1, X_4, and X_5. By doing this, the probability of loss caused by fund-raising fraud can be reduced.

3.4. *Impact Analysis of Safety Measures*

3.4.1. The Effect of Fraud Rate on Fraud Dissemination

To study the impact of the fraud rate of fund-raising fraud on fraud dissemination, a simulation analysis is carried out by changing the fraud rate. The fraud rate β is set as 0.25, 0.2, 0.15, and 0.1 for simulation, and the simulation results are shown in Figure 10.

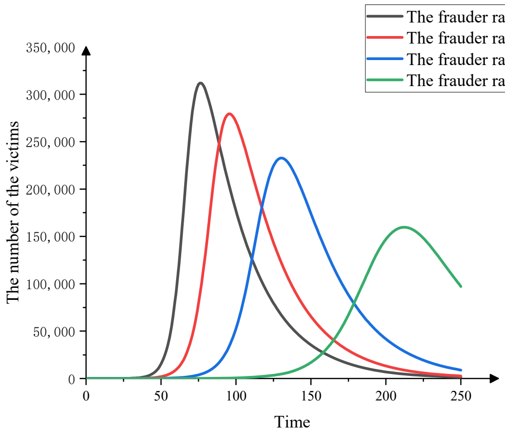


Figure 10. Impact of reduced fraud rates on fraud dissemination.

According to the analysis in Figure 10, when β = 0.25, the number of maximum victims reaches 310,000, and the arrival time of the maximum victims is 75 days; when β = 0.2, the number of maximum victims reaches 278,000, and the arrival time of the maximum victims is 97 days; when β = 0.15, the number of maximum victims reaches 232,000, and the arrival time of the maximum victims is 130 days; when β = 0.1, the number of maximum victims reaches 159,000, and the arrival time of the maximum victims is 212 days. With the decrease in the fraud rate β, the arrival time of the maximum victims is prolonged, and the number of maximum victims gradually decreases. Therefore, it can be concluded that reducing the fraud rate can reduce the number of victims and delay the outbreak time of the fraud.

3.4.2. Alert Rate Effect on the Spread of the Fraud

To study the impact of the alert rate of fund-raising fraud on fraud dissemination, simulation analysis is carried out by changing the alert rate of fraud. The alert rate γ is set as 0.032, 0.082, 0.132, and 0.182 for simulation. The simulation results are shown in Figure 11.

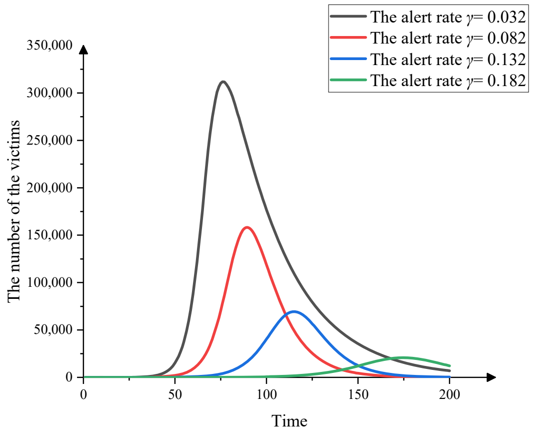


Figure 11. Impact of increased alert rates on fraud dissemination.

According to the analysis in Figure 11, when γ = 0.032, the number of maximum victims reaches 310,000, and the arrival time of the maximum victims is 75 days; when γ = 0.082, the number of maximum victims reaches 158,000, and the arrival time of the maximum victims is 89 days; when γ = 0.132, the number of maximum victims reaches 69,000, and the arrival time of the maximum victims is 115 days; when γ = 0.182, the number of maximum victims reaches 20,000, and the arrival time of the maximum victims is 175 days. With the increase of the alert rate γ, the arrival time of the maximum victims is also prolonged, and the number of maximum victims gradually decreases. From the above analysis, it can be seen that improving the alert rate can greatly reduce the number of victims and effectively curb the dissemination of fraud.

3.4.3. The Effect of Fraud Supervision on Fraud Dissemination

To study the influence of measures to strengthen fraud supervision of fraud dissemination, the regulation coefficient $\boldsymbol{K}$ is introduced to modify the fraud rate for the simulation analysis [42], as shown in Equations (17) and (18).

$$\boldsymbol{K = e^{-mI(t)}} \tag{17}$$

$$\boldsymbol{\beta' = K\beta,} \tag{18}$$

where $\boldsymbol{m}$ represents the regulatory impact factor, and $\boldsymbol{\beta'}$ represents the fraud rate of fund-raising fraud under the influence of regulatory authorities.

The regulatory impact factor $\boldsymbol{m}$ is set as 5×10^{-5}, 1×10^{-4}, 1.5×10^{-4}, and 2×10^{-4}, respectively for the simulation, and the simulation results are shown in Figure 12.

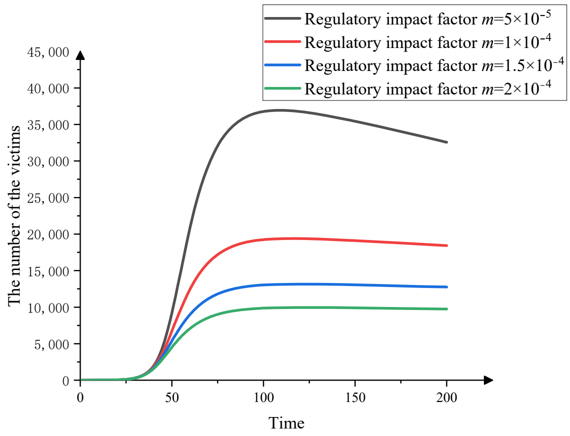


Figure 12. Impact of regulatory impact factor on fraud dissemination.

According to the analysis in Figure 12, when $\boldsymbol{m} = 5 \times 10^{-5}$, the number of maximum victims reaches 37,000, and the arrival time of the maximum victims is 109 days; when $\boldsymbol{m} = 1 \times 10^{-4}$, the number of maximum victims reaches 19,400, and the arrival time of the maximum victims is 115 days; when $\boldsymbol{m} = 1.5 \times 10^{-4}$, the number of maximum victims reaches 13,100, and the arrival time of the maximum victims is 119 days; when $\boldsymbol{m} = 2 \times 10^{-4}$, the number of maximum victims reaches 995,000, and the arrival time of the maximum victims is 115 days. With the increase of the regulatory impact factor $\boldsymbol{m}$, swindlers avoid the impact of regulation by reducing the scale of fund-raising fraud. The number of maximum victims is greatly reduced, and the number of the new victims is reduced, but the time for victims to return to alerts is greatly prolonged.

4. Conclusions

In this study, the SIR model, BP neural network, Fault tree, and Bayesian network were used to analyze fund-raising fraud. Firstly, relevant data were collected from a questionnaire survey. Secondly, the SIR model was used to simulate the dissemination; the BP neural network was used to identify the data; the Fault tree and the Bayesian network were employed to analyze the causation and importance of basic events. Finally, the security measures of fund-raising fraud were simulated by changing the dissemination parameters. The results show that:

(1) Numerical simulation of short-term fund-raising fraud based on the SIR model

In the first fraud cycle, there is a low fraud rate. At this stage, the dissemination degree of fund-raising fraud is low, and it is difficult for it to be detected by the regulatory authorities due to the small number of victims. In the second fraud cycle, the monthly investment profit results in an increase in the fraud rate and a decreasing alert rate. A large scale can be formed. In the third fraud cycle, due to the returned profits and publicity, the fraud rate increases, the alert rate decreases, and an influenza-type fraud dissemination can be formed. As a result, huge losses can be caused, and the victims of the final fraud cycle account for 12.5% of people in the region. Therefore, fraud dissemination must be stopped in the early development stages.

(2) Recognition of fraud data based on BP neural network

The recognition accuracy of the test set data in the BP neural network reaches 90.9%, indicating that fund-raising fraud can be preliminarily identified through the BP neural network. With the increase in operation step size, the recognition rate changes from 90.9% to 93.9%, and the error rate decreases gradually. When the step sizes are 500 and 600, the error rate is stable at 4.5%. With the increase in the number of hidden neurons, the recognition rate remains stable at 90.9%, and the error rate decreases from 6% to 3%.

(3) System analysis based on fault tree and Bayesian networks

According to the calculation result of the Bayesian network, although the probability of loss caused by fraud is only 0.0498, huge losses can be caused by the increase of fraud dissemination and fraud amount. Through the analysis of the structural importance, probability importance, and critical importance, the occurrence of basic events $\boldsymbol{X_1}$, $\boldsymbol{X_4}$, and $\boldsymbol{X_5}$ has a significant impact on the loss caused by fund-raising fraud. Relevant measures can be taken to reduce the probability of occurrence of $\boldsymbol{X_1}$, $\boldsymbol{X_4}$, and $\boldsymbol{X_5}$, and, thus, reduce the probability of loss caused by fund-raising fraud.

(4) Impact of safety measures

The results of the simulation analysis show that decreasing the fraud rate can reduce some victims and delay the outbreak of the fraud, increasing the alert rate can reduce the number of victims on a large scale, and strengthening the regulation can greatly restrain the scale of victims but extend the duration of a fraud.

Author Contributions: Conceptualization, S.Y.; Methodology, S.Y.; Data curation, K.S. and Z.X.; Formal analysis, K.S. and Z.X.; Validation, S.Y.; Resources, S.Y. and K.S.; Writing—original draft preparation, K.S.; Writing—review and editing, K.S.; Project administration, S.Y. and B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hunan Provincial Natural Science Foundation of China under Grant No. 2021JJ40799 and the Education reform Project of Central South University under Grant No. 2021JGB095.

Institutional Review Board Statement: It is our understanding that our research was conducted following the ethics requirements of our universities and is exempt from research ethics committee oversight. The reason is that the subjects in our survey cannot be identified in any way or expose any risks, liabilities, or reputational damage.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is contained within the article.

Acknowledgments: The authors would like to express their thanks to the Hunan Provincial Natural Science Foundation and Education reform Project of Central South University.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. The questionnaire survey project description table.

	Questionnaire
1	Age
	A. 21–30 years old B. 31–40 years old C. 41–50 years old D. 41–50 years old E. Above 50 years old
2	Gender
	A. Female B. Male
3	Education level
	A. Primary school B. Junior high school C. Senior high school D. University E. Master F. Doctor
4	Would you like to participate in the investment? (such as funds, stocks, real estate, bonds, precious metals, etc.)
	A. Yes B. No
5	Have you received messages of investment promotion with high investment and high profit?
	A. Yes B. No
6	How do you make investment decisions?
	A. Independent choice B. Others suggest
7	Would you invest in products with the following monthly interest rates?
	A. 1% B. 5% C. 10% D. 20% E. 40%
8	Do you have professional investment knowledge when investing in financial products?
	A. Yes B. No
9	Have you received any warning from fraud prevention agencies?
	A. Yes B. No

Table A1. *Cont.*

	Questionnaire
10	How much money do you choose to invest in financial products for the first time? (Please give an amount from the range below)
	A. 2000–4000 B. 4000–6000 C. Above 6000
11	Would you withdraw your investment within the first month?
	A. Yes B. No
12	Would you continue to invest in the following wealth management products after receiving the first month's investment income?
	A. Yes B. No
13	How much money do you choose to invest in financial products for the second time? (Please give an amount from the range below)
	A. 6000–8000 B. 8000–10,000 C. Above 10,000
14	Would you withdraw your investment within the second month?
	A. Yes B. No
15	Would you continue to invest in the following wealth management products after receiving the investment income of the second month?
	A. Yes B. No
16	How much money do you choose to invest in financial products for the third time? (Please give an amount from the range below)
	A. 10,000–12,000 B. 12,000–14,000 C. Above 14,000
17	Would you withdraw your investment within the third month?
	A. Yes B. No

Appendix B

Table A2. The descriptive statistics of corresponding variables.

Variables	**Total (n = 1032)**	**Losses Caused (n = 258.25%)**	**No Losses Caused (n = 774.75%)**	χ^2	p
		Age			
21–30 years old	103 (9.9)	31 (3)	72 (6.9)	2.763	0.167
31–40 years old	412 (39.9)	97 (9.3)	315 (30.5)		
41–50 years old	468 (45.3)	105 (10.1)	363 (35.1)		
Above 50 years old	49 (4.7)	12 (1.1)	37 (3.6)		
		Gender			
Female	412 (39.9)	91 (8.8)	321 (31.1)	3.103	0.077
Male	620 (60.1)	167 (16.1)	453 (43.8)		
		Education level			
Primary school	62 (6)	16 (1.5)	46 (4.4)	0.028	0.273
Junior high school	206 (19.9)	51 (4.9)	155 (15)		
Senior high school	124 (12)	31 (3)	93 (9)		
University	616 (59.6)	154 (14.9)	462 (44.7)		
Master	20 ((1.9)	5 (0.4)	15 (1.4)		
Doctor	4 (0.3)	1 (0.1)	3 (0.2)		

Table A2. *Cont.*

Variables	Total (n = 1032)	Losses Caused (n = 258.25%)	No Losses Caused (n = 774.75%)	χ^2	p
		Participate in the investment			
Yes	825 (79.9)	207 (20)	618 (59.9)	0.018	0.4
No	207 (20)	51 (4.9)	156 (15.1)		
		Receive messages of investment			
Yes	145 (14)	131 (12.6)	14 (1.4)	384.189	0.56
No	887 (85.9)	127 (12.3)	760 (73.6)		
		Make investment decisions			
Independent choice	186 (18)	53 (5.1)	133 (12.8)	1.478	0.433
Others suggest	846 (81.9)	205 (19.8)	641 (62.1)		
		Monthly interest rates			
1%	31 (3)	8 (0.7)	23 (2.2)	0.044	0.152
5%	154 (14.9)	39 (3.7)	115 (11.1)		
10%	208 (20.1)	52 (5)	156 (15.1)		
20%	516 (50)	129 (12.5)	387 (37.5)		
40%	123 (11.9)	30 (2.9)	93 (9)		
		Lack of professional investment knowledge			
Yes	413 (40)	103 (9.9)	310 (30)	0.001	0.086
No	619 (59.9)	155 (15)	464 (44.9)		
		Receive warning from fraud prevention agencies			
Yes	213 (20.6)	54 (5.2)	159 (15.4)	0.004	0.396
No	819 (79.3)	206 (19.9)	613 (59.3)		
		Amount of the first investment			
2000–4000	920 (89.1)	230 (22.2)	690 (66.8)	4.14	0.512
4000–6000	112 (10.8)	38 (3.6)	74 (7.1)		
Above 6000	0	0	0		
		Withdraw your investment within the first month			
Yes	66 (6.3)	17 (1.6)	49 (4.7)	0.022	0.544
No	966 (93.6)	241 (23.3)	725 (70.2)		
		Continue to invest in products after receiving investment income			
Yes	170 (16.4)	43 (4.1)	127 (12.3)	0.009	0.445
No	862 (83.5)	215 (20.8)	647 (62.6)		
		Amount of the second investment			
6000–8000	620 (60)	158 (15.3)	462 (44.7)	0.194	0.297
8000–10,000	412 (39.9)	100 (9.6)	312 (30.2)		
Above 10,000	0	0	0		
		Withdraw your investment within the second month			
Yes	58 (5.6)	15 (1.4)	43 (4.1)	0.024	0.55
No	974 (94.3)	243 (23.5)	731 (70.8)		
		Continue to invest products after receiving investment income again			
Yes	258 (25)	62 (6)	196 (19)	0.172	0.33
No	774 (75)	196 (19)	578 (56)		

Table A2. *Cont.*

Variables	Total (n = 1032)	Losses Caused (n = 258.25%)	No Losses Caused (n = 774.75%)	χ^2	p
		Amount of the third investment			
10,000–12,000	151 (15.3)	38 (3.6)	113 (10.9)	0.035	0.429
12,000–14,000	835 (80.9)	208 (20.1)	627 (60.7)		
Above 14,000	46 (4.4)	12 (1.1)	34 (3.2)		
		Withdraw your investment within the third month			
Yes	33 (3.2)	0	33 (3.2)	11.363	0.564
No	999 (96.8)	258 (25)	741 (71.8)		

Appendix C

The list of control variables used in the analysis:

$\boldsymbol{\beta_1}$ is the fraud rate in the first fraud cycle. The data comes from the frequency of Yes to question 5 on the questionnaire.

$\boldsymbol{\beta_2}$ is the fraud rate in the second fraud cycle. The data comes from the frequency of Yes to question 12 on the questionnaire.

$\boldsymbol{\beta_3}$ is the fraud rate in the third fraud cycle. The data comes from the frequency of Yes to question 15 on the questionnaire.

γ_1 is the alert rate in the first fraud cycle. The data comes from the frequency of No to question 11 on the questionnaire.

γ_2 is the alert rate in the first fraud cycle. The data comes from the frequency of No to question 14 on the questionnaire.

γ_3 is the alert rate in the first fraud cycle. The data comes from the frequency of No to question 17 on the questionnaire.

$\boldsymbol{C_1}$ is the average investment amount of the first product. The data comes from the average amount of money answered in question 10.

$\boldsymbol{C_2}$ is the average investment amount of the second product. The data comes from the average amount of money answered in question 13.

$\boldsymbol{C_3}$ is the average investment amount of the third product. The data comes from the average amount of money answered in question 16.

$\boldsymbol{q_1}$ is the probability of $\boldsymbol{X_1}$. The data comes from the frequency of Yes to question 5 on the questionnaire.

$\boldsymbol{q_2}$ is the probability of $\boldsymbol{X_2}$. The data comes from the frequency of No to question 8 on the questionnaire.

$\boldsymbol{q_3}$ is the probability of $\boldsymbol{X_3}$. The data comes from the frequency of "Others suggest" to question 6 on the questionnaire.

$\boldsymbol{q_4}$ is the probability of $\boldsymbol{X_4}$. The data comes from the frequency of No to question 9 on the questionnaire.

$\boldsymbol{q_5}$ is the probability of $\boldsymbol{X_5}$. The data comes from the frequency of No to question 11 on the questionnaire.

Table A3. Statistical results SIR parameter table.

Related Parameters	β_i	C_i	γ_i	A
$i = 1$	0.131	3020	0.064	20%
$i = 2$	0.164	8100	0.055	20%
$i = 3$	0.25	13,200	0.032	20%

Table A4. Training set data.

Number	Investment Amount (CNY)	Monthly Interest Rate	Class	Number	Investment Amount	Monthly Interest Rate	Class
1	2201	40%	1	...	...	...	...
2	3551	40%	1	604	3326	10%	2
3	3634	40%	1	605	2893	10%	2
4	2450	40%	1	606	3753	10%	1
5	3144	40%	1	607	3200	10%	2
6	3594	40%	1	608	2957	10%	2
7	3379	1%	2	609	3119	10%	2
8	2324	1%	2	610	2925	10%	2
9	3384	1%	2	612	2822	10%	2
10	3658	1%	2	613	2952	20%	2
11	2759	1%	2	614	3692	20%	1
12	2043	5%	2	615	2229	20%	2
13	3346	5%	2	616	3788	20%	1
14	3366	5%	2	617	2367	20%	2
15	2918	5%	2	618	3581	20%	1
...	...	...	...	619	3834	20%	1

Table A5. Test set data.

Number	Investment Amount (CNY)	Monthly Interest Rate	Class	Number	Investment Amount	Monthly Interest Rate	Class
1	3290	20%	1	...	...	...	...
2	2168	20%	2	399	2066	20%	2
3	2952	20%	2	400	3492	20%	1
4	3670	20%	1	401	2768	20%	2
5	3144	20%	2	402	2562	20%	2
6	2270	20%	2	403	3302	20%	2
7	2785	20%	2	404	2886	20%	2
8	2270	20%	2	405	2029	20%	2
9	2155	20%	2	406	2133	20%	2
10	3180	20%	2	407	2430	20%	2
11	3501	20%	1	408	2912	20%	2
12	3578	20%	1	409	2556	20%	2
13	2193	20%	2	410	3547	20%	1
14	2505	20%	2	411	2241	20%	2
15	3831	20%	1	412	3999	20%	1
...	...	...	...	413	2518	20%	2

Table A6. Event name and base probability.

Event Number	Event Name	Event Probability
X_1	Receiving fraudulent promotional messages	0.13
X_2	Lack of professional knowledge	0.4
X_3	Over trust others' recommendation	0.18
X_4	Lack of regulation of financial fraud	0.8
X_5	Lack of risk awareness after investment	0.064

Table A7. Conditional probability distribution of M_2.

X_3	X_2	Yes	No
Yes	Yes	29%	71%
Yes	No	18%	82%
No	Yes	40%	60%
No	No	0	100%

Table A8. Conditional probability distribution of M_1.

M_2	X_4	X_1	Yes	No
Yes	Yes	Yes	18.3%	81.7%
Yes	Yes	No	20.9%	79.1%
Yes	No	Yes	17.3%	82.7%
Yes	No	No	21.7%	78.3%
No	Yes	Yes	16.5%	83.5%
No	Yes	No	20%	80%
No	No	Yes	13%	87%
No	No	No	0	100%

Table A9. Conditional probability distribution of T.

M_1	X_5	X_4	Yes	No
Yes	Yes	Yes	14.1%	85.9%
Yes	Yes	No	18%	82%
Yes	No	Yes	11.2%	88.8%
Yes	No	No	16%	84%
No	Yes	Yes	13.2%	86.8%
No	Yes	No	20%	60%
No	No	Yes	6.4%	93.6%
No	No	No	0	100%

References

1. Bartoletti, M.; Pes, B.; Serusi, S. Data Mining for Detecting Bitcoin Ponzi Schemes. In Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), Zug, Switzerland, 20–22 June 2018; pp. 75–84. [CrossRef]
2. Yu, W.Q.; Zhang, Y.M.; Li, Z.Y.; Niu, W. Type analysis and identification method of Ethereum Ponzi scheme. *J. Chongqing Univ.* **2020**, *43*, 111–120. [CrossRef]
3. Zhou, Y.C. Research on Bitcoin Ponzi Scheme Detection Based on Data Mining Technology. Master's Thesis, Guangdong University of Technology, Guangzhou, China, 2020. [CrossRef]
4. Zhang, Y.M.; Lou, Y.C. Ponzi scheme contract detection method based on deep neural network. *Comput. Sci.* **2021**, *48*, 273–279. [CrossRef]
5. Bayraktar, E.; Cohen, A.; Nellis, A. A Macroeconomic SIR Model for COVID-19. *Mathematics* **2021**, *9*, 1901. [CrossRef]
6. Liu, C.; Han, R.; Huang, X.Y.; Yang, H.Y.; Liu, X.Y. The Internet pyramid selling SIR propagation model. *J. Chongqing Univ. Technol.* **2021**, *35*, 161–167.
7. Klafft, M. *Peer to Peer Lending: Auctioning Microcredits over the Internet*; Social Science Electronic Publishing: Rochester, NY, USA, 2008; Available online: https://ssrn.com/abstract=1352383 (accessed on 1 April 2022).
8. Vasek, M.; Moore, T. *Analyzing the Bitcoin Ponzi Scheme Ecosystem*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 101–112. [CrossRef]
9. Slattery, P.D. Square Pegs in a Round Hole: SEC Regulation of Online Peer-to-Peer Lending and the CFPB Alternative. *Yale J. Regul.* **2013**, *30*, 6. Available online: https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=1369&context=yjreg&httpsredir=1&referer= (accessed on 1 April 2022).
10. Fan, J.X. Research on Countermeasures to Prevent Online Financial Fraud Crime. Master's Thesis, Jiangxi University of Finance and Economics, Nanchang, China, 2017. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=1017206147.nh&DbName=CMFD2018 (accessed on 1 April 2022).
11. Fan, X. Investigation Report on Baoding Citizens Encountering Financial Fraud. Master's Thesis, Hebei University of Finance and Economics, Shijiazhuang, China, 2017. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=1017020187.nh&DbName=CMFD2018 (accessed on 1 April 2022).
12. Kermack, W.; McKendrick, A. Contributions to the mathematical theory of epidemics—I. *Bull. Math. Biol.* **1991**, *53*, 33–55. [CrossRef] [PubMed]
13. Jayatilaka, R.; Patel, R.; Brar, M.; Tang, Y.; Jisrawi, N.; Chishtie, F.; Drozd, J.; Valluri, S. A mathematical model of COVID-19 transmission. *Mater. Today Proc.* **2022**, *54*, 101–112. [CrossRef]
14. Razaque, A.; Rizvi, S.; Khan, M.J.; Almiani, M.; Al Rahayfeh, A. State-of-art review of information diffusion models and their impact on social network vulnerabilities. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 1275–1294. [CrossRef]
15. Ma, X.; Deng, W.; Qiao, W.; Lan, H. A methodology to quantify the risk propagation of hazardous events for ship grounding accidents based on directed CN. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108334. [CrossRef]
16. Gong, J.; Williams, M.A.; McAfee, R.P. Fraud Cycles. *J. Inst. Theor. Econ.* **2016**, *172*, 544–572. [CrossRef]

17. Wang, Y.; Stuart, T.; Li, J. Fraud and Innovation. *Adm. Sci. Q.* **2021**, *66*, 267–297. [CrossRef]
18. Bhattacharya, U. The optimal design of Ponzi schemes in finite economies. *J. Financ. Intermediation* **2003**, *12*, 2–24. [CrossRef]
19. Xu, L.; Wang, J.; Xu, D.; Xu, L. Integrating Individual Factors to Construct Recognition Models of Consumer Fraud Victimization. *Int. J. Environ. Res. Public Healh* **2022**, *19*, 461. [CrossRef] [PubMed]
20. Cuijie, Z. Research of expression recognition base on optimized BP neural network. In Proceedings of the 2009 16th International Conference on Industrial Engineering and Engineering Management, Beijing, China, 21–23 October 2009; pp. 1803–1806. [CrossRef]
21. Lu, Y.; Li, Z.; Zhao, X.; Lv, S.; Wang, X.; Wang, K.; Ni, H. Recognition of Rice Sheath Blight Based on a Backpropagation Neural Network. *Electronics* **2021**, *10*, 2907. [CrossRef]
22. Wang, Y.; Chen, C.; Yan, X. Structure and weight optimization of neural network based on CPA-MLR and its application in naphtha dry point soft sensor. *Neural Comput. Appl.* **2013**, *22*, 75–82. [CrossRef]
23. Ronza, A.; Félez, S.; Darbra, R.M.; Carol, S.; Vílchez, J.; Casal, J. Predicting the frequency of accidents in port areas by developing event trees from historical analysis. *J. Loss Prev. Process Ind.* **2003**, *16*, 551–560. [CrossRef]
24. Fang, M.; Zhang, Y.; Zhu, M.; Chen, S. Cause Mechanism of Metro Collapse Accident Based on Risk Coupling. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2102. [CrossRef] [PubMed]
25. Özbay, C.; Özbay, T.; Yiğitoğlu, A.G.; Bayburt, M. Probabilistic risk assessment of radiotherapy application. *Radioprotection* **2022**, *57*, 33–40. [CrossRef]
26. Barua, S.; Gao, X.; Pasman, H.; Mannan, M.S. Bayesian network based dynamic operational risk assessment. *J. Loss Prev. Process Ind.* **2016**, *41*, 399–410. [CrossRef]
27. Lalika, L.; Kitali, A.E.; Haule, H.J.; Kidando, E.; Sando, T.; Alluri, P. What are the leading causes of fatal and severe injury crashes involving older pedestrian? Evidence from Bayesian network model. *J. Saf. Res.* **2022**, *80*, 281–292. [CrossRef]
28. Hunte, J.L.; Neil, M.; Fenton, N.E. A causal Bayesian network approach for consumer product safety and risk assessment. *J. Saf. Res.* **2022**, *80*, 198–214. [CrossRef] [PubMed]
29. Mahadevan, S.; Zhang, R.; Smith, N. Bayesian networks for system reliability reassessment. *Struct. Saf.* **2001**, *23*, 231–251. [CrossRef]
30. Liu, Y.; Ma, X.; Qiao, W.; Luo, H.; He, P. Human Factor Risk Modeling for Shipyard Operation by Mapping Fuzzy Fault Tree into Bayesian Network. *Int. J. Environ. Res. Public Health* **2022**, *19*, 297. [CrossRef] [PubMed]
31. Xin, S.; Zhang, L.; Jin, X.; Zhang, Q. Reconstruction of the fault tree based on accident evolution. *Process Saf. Environ. Prot.* **2019**, *121*, 307–311. [CrossRef]
32. Wei, C.R.; Sun, J.H.; Zhang, J.P. Qualitative analysis of accident tree and its application in mine safety evaluation. *Industrial Saf. Environ. Prot.* **2009**, *35*, 39–41. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=GYAF200909017&DbName=CJFQ2009 (accessed on 1 April 2022).
33. Li, Y.X.; Sun, J.H.; Wei, C.R. Research and discussion on solving method of importance degree of the accident tree structure. *China Saf. Sci. Technol.* **2012**, *8*, 107–110. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=LDBK201205021&DbName=CJFQ2012 (accessed on 1 April 2022).
34. Zaib, A.; Yin, J.; Khan, R.U. Determining Role of Human Factors in Maritime Transportation Accidents by Fuzzy Fault Tree Analysis (FFTA). *J. Mar. Sci. Eng.* **2022**, *10*, 381. [CrossRef]
35. Xi, J.; Zhao, Y.; Ding, T.; Tian, J.; Li, L. Analysis Model of Risk Factors of Urban Bus Operation Based on FTA-CLR. *Adv. Civ. Eng.* **2021**, *2021*, 6657786. [CrossRef]
36. Xin, S.; Zhu, X.; Liu, S.; Guo, J. Research on Fault Tree Reconstruction Based on Contingency. *Processes* **2022**, *10*, 427. [CrossRef]
37. Sheng, B.; Deng, C.; Wang, Y.H.; Tang, L.H. System Analysis by Mapping a Fault-tree into a Bayesian-network. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *362*, 012025. [CrossRef]
38. Del-Valle, M.V.; López-Morales, H.; Andrés, M.L.; Yerro-Avincetto, M.; Trudo, R.G.; Urquijo, S.; Canet-Juric, L. Intolerance of COVID-19-related uncertainty and depressive and anxiety symptoms during the pandemic: A longitudinal study in Argentina. *J. Anxiety Disord.* **2022**, *86*, 102531. [CrossRef] [PubMed]
39. McHugh, M.L. The Chi-square test of independence. *Biochem. Med.* **2013**, *23*, 143–149. [CrossRef] [PubMed]
40. Zang, X. Review and Identification of Evidence in Internet Financial Crimes-A Case study of Ezubao. *Law Soc.* **2021**, 24–26. [CrossRef]
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
42. Huo, L.; Wang, L.; Song, N.; Ma, C.; He, B. Rumor spreading model considering the activity of spreaders in the homogeneous network. *Phys. A Stat. Mech. Appl.* **2017**, *468*, 855–865. [CrossRef]