Research Article

Judgment of Social Governance System Risk Based on Big Data Analysis

Xiaojing Yang ⬤

Hangzhou Youth & Children's Center, Zhejiang, Hangzhou 310007, China

Correspondence should be addressed to Xiaojing Yang; 631605040225@mails.cqjtu.edu.cn

Accurate risk identification, scientific risk assessment, effective risk early warning, and real-time risk monitoring will all benefit from the advancement of data science and the development of the data industry. Simultaneously, social risk prevention and control face numerous obstacles in the growth of information technology, data integration and mining, data information disclosure, and data culture popularization in the era of big data. It can be improved in terms of risk prevention and control concept innovation. In addition, it can also be improved in terms of network public opinion guidance, control mechanism improvement, social security prevention, control system optimization, and the introduction and training of big data talent to promote scientific and accurate social risk prevention and control. As society moves into the 5.0 era, many social risks rise and new features emerge, posing larger threats and unprecedented challenges to social risk governance. Big data, as a result of the emerging technology revolution, offers a wealth of governance resources and brand-new governance concepts that traditional governance mechanisms lack, offering technical assistance. This article presents a parallel community partition algorithm using MapReduce to analyze big data. The proposed model is established for the big data thinking paradigm to promote social risk governance innovation. The proposed model supports the top-level model and does inclusive forecasting for social risk management that supports resource integration. The proposed model also realizes diversified and coordinated risk governance. The results reveal the significance of the proposed model.

1. Introduction

Today's world is undergoing massive transformations that have not been seen in a century [1]. A new cycle of scientific and technical revolutions is in full swing. The international power balance is being rebalanced, and the vision of forming a community with a common future for humanity is firmly rooted in the hearts of the people. Simultaneously, the global environment has become increasingly complex, instability and uncertainty have increased significantly, and the impact of the COVID-19 epidemic has been widespread and far-reaching [2]. The global economy has entered a downturn, economic globalization has run into a countercurrent, and the world has entered a new era. Economic globalization is the general trend, the characteristics of the world's multi-polarization pattern are clear, science, technology, and industry reform are extensive, and "development and security" remain the two major themes of today's era's development.

China will initiate a novel pathway of constructing an up-to-date socialist republic in all aspects and marching towards the second centennial goal after finishing the first centennial objective of constructing a well-off society in all aspects and achieving the 14th Five Year Plan. The requirements of "coordinating development and security and building a higher level of safe China," "further strengthening the national security system and capacity-building," "further strengthening national economic security," "comprehensively improving public security capacity," and "maintaining social stability and security" have all been clearly stated by China. China will have completed socialist modernization in the year 2035 [3]. Economic strength, scientific and technological strength, and overall national strength will all increase dramatically, total economic output and per capita income of urban and rural residents will both increase significantly, and breakthroughs in key core technologies will be made, propelling the country into the forefront of innovative countries. Create a

contemporary economic structure by implementing new industrialization, information technology, urbanization, and agricultural modernization.

The background of the era of big data is the explosive growth of data [4]. Big data plays a vital role in the current era. Big data is the backbone of analytics and prediction. Big data analytics platforms are used to analyze big data in a parallel fashion [5–7]. The parallel and distributed platforms are used to compute the big data. At the beginning of the twenty-first century, the amount of data available around the world is about several EBs (1EB = 1018 B) or tens of thousands of times the total number of all books in the world. At present, we are in the transition from PC Internet to mobile Internet, and mobile phones and other mobile devices also produce a large amount of data. Human genome engineering reveals complex interactions between billions of cells, proteins, and genes, which will generate Pangian bioinformatic data. The amount of data generated worldwide has been increased by several orders of magnitude to the ZB level (1 ZB = 1021 B). More data will be generated in the future. By 2020, the data volume is expected to be more than ten times higher [8–10]. First and foremost, we must create a method for sharing information. Government departments should speed up the elimination of institutional barriers and break the dilemma of information obstruction under the restrictions of traditional levels. Governments at all levels should rely on the information support of big data, break the separation of interests and "small peasant consciousness," speed up the interaction between data, formulate unified data collection and storage standards, and improve the government's basic information database, especially in areas close to people's lives.

Through big data technology, integrate and optimize the business processes of various departments, establish a strong information exchange, seamless connection, communication, and coordination system, form a standardized cooperative governance framework, accurately reverse the offside and dislocation problems of government departments, improve the horizontal consistency of departments, and jointly deal with complex social governance problems. Third, we should improve the prediction and early warning mechanism. Developing the in-depth value of big data and establishing an emergency early warning mechanism covering the whole area of social governance is an important choice to improve the national security system. It is necessary to organically connect all kinds of governance data through the collection, sorting, and calculation functions of big data, establish a composite prediction model, and monitor the social governance status in real time. According to the analysis, classification, hierarchical analysis, and visual display of big data on the corresponding index data, set up a multigradient response mode according to the progress of the situation and continuously improve the ability of government departments to deal with various public crisis events.

The requirements for data processing capabilities are also becoming higher and higher with the generation of human quantity data. Human data processing mode mainly includes batch processing mode and flow processing mode. Batch

human data analysis adopts batch processing mode, first data storage, and then the static storage data set. At present, the most common mass human data computing framework is Hadoop [11–13], which mainly includes HDFS (Hadoop distributed file system) and MapReduce [14–16], which are responsible for storing static data, while the latter completes the allocation of computing tasks. Each data node gets the corresponding computing tasks, respectively, and all nodes cooperate for data analysis and value discovery.

2. Related Works

The literature in the context of the research and judgment of social governance system risk based on big data analysis is provided in this section. The literature is discussed in the context of the main stages of social governance development and online learning algorithms overview.

2.1. Three Main Stages of Social Governance Development. The party and the state have always placed a high priority on social management since the formation of the new China. They have conducted extensive research and practice to establish and develop a social management system that is appropriate for China's national circumstances, and they have achieved major results and gained considerable expertise. "Comprehensive management centers" have been established in a few locations during the 1990s [17, 18]. It has taken almost 30 years up to this point. We have complied with the more diverse development of socioeconomic components particularly since reform and opening up and have continually supported social management reform and innovation. From the Third Plenary Session of the 14th CPC Central Committee proposing to strengthen the government's social management function to the Fourth Plenary Session of the 16th CPC Central Committee proposing to strengthen social construction and management to the Sixth Plenary Session of the 16th CPC Central Committee emphasizing the innovation of social management system and the integration of social management resources to the 17th CPC National Congress emphasizing the improvement of social management pattern and grass-roots social management system, the CPC's understanding of social management has been deepened.

China has established a leadership system for social management in the long-term exploration and practice. China also built a social management organization network and formulated basic laws and regulations for social management. China continuously promoted the adaptation of social management to China's national conditions and the socialist system [19]. Although it is a one-word difference, it is a comprehensive improvement of the party's ruling philosophy and policy ideas in the social field, which reflects the systematic governance at the source and comprehensive implementation of policies. Social management puts more emphasis on the unitary leadership of the government and it emphasizes multiple participation. In the development process of "social governance," the "comprehensive management center" has been playing an important role for

more than 30 years, and its development has roughly experienced three important historical stages.

The task is to implement comprehensive management of social security, which is a significant facet of communal authority. What we pay attention to and solve is the problem of social security. The second decade is comprehensive social management. The task is to implement comprehensive management of social management, which has expanded the extension based on comprehensive management of social security but has not yet risen to the level of social governance. The third decade is comprehensive social governance.

2.2. Overview of Online Learning Algorithms. The online learning algorithm includes several approaches and models. One of the most widely used models is the perceptron [20–22]. Perceptron is a machine learning bionics model that consists of two classification learning machines. The classification learning machines are used to classify or categorize the outcome based on the specific training. The classifiers are trained using the labeled data. The labels are the actual outputs against certain inputs. Many complex algorithms are based on Rosenblatt's perceptron algorithm. When the classifiers are correct, the $\overrightarrow{w}$ (weight vector) is "rewarded"; that is, it remains untouched; when the classification is incorrect, the weight vector $\overrightarrow{w}$ is "fined"; that is, it is adjusted and transformed to the correct direction. For the weight vector $\overrightarrow{w}$, if a sample feature vector $\overrightarrow{x}_i$ is misclassified, then $\overrightarrow{w}^T \overrightarrow{x}_i \leq 0$. To express the punishment for misdivided samples, we sum the mode functions of all misdivided samples, as shown in

$$J_P(\overrightarrow{w}) = \sum_{i \in \Gamma} \left(-\overrightarrow{w}^T \overrightarrow{x}_i \right), \tag{1}$$

where Γ is the subscript set of the samples misclassified by the hyperplane of equation (1), and $J_P(\overrightarrow{w})$ is the risk functional, if and only if a specific condition related to weight is met. The specific condition 2 is expressed as formula (2). The weight $\overrightarrow{w}^*$ is the solution vector.

$$J_P(\overrightarrow{w}^*) = \min J_P(\overrightarrow{w}) = 0. \tag{2}$$

The minimization of the perceptron criterion function can be solved by the gradient descent method described as follows:

$$\overrightarrow{w}_t = \overrightarrow{w}_{t-1} - \eta \nabla J_P(\overrightarrow{w}), \tag{3}$$

where t is the number of iterations and $\eta > 0$ is the adjusted step size. According to equation (1),

$$\nabla J_P(\overrightarrow{w}) = \frac{\partial J_P(\overrightarrow{w})}{\partial \overrightarrow{w}},$$
$$= \sum_{i \in \Gamma} (-\overrightarrow{x}_i). \tag{4}$$

So, equation (3) becomes

$$\overrightarrow{w}_t = \overrightarrow{w}_{t-1} - \eta,$$
$$= \sum_{i \in \Gamma} (-\overrightarrow{x}_i). \tag{5}$$

That is, the misclassified samples are superimposed on the weight vector according to a coefficient in each iteration. The misclassified samples may not be considered for the comparative analysis of the confusion matrix. The training testing ratio is kept at 70–30 in the classification algorithm.

The steps of the perceptron algorithm are as follows:

(1) To create a training sample set, select n pattern samples from both positive and negative categories, as stated in equation (1). Multiply all negative samples by (–1). $\overrightarrow{w}_0$ is the initial value of any weight vector, and the iteration starts at $t = 1$.

(2) Calculate the $\overrightarrow{w}_{t-1}^T \overrightarrow{x}_i$ value of all training samples, carry out a round of iteration, and correct the weight vector.

$$\overrightarrow{w}_t = \begin{cases} \overrightarrow{w}_{t-1}, & \overrightarrow{w}_{t-1}^T \overrightarrow{x}_i \geq 0, \\ \overrightarrow{w}_{t-1} + \eta \overrightarrow{x}_i, & \overrightarrow{w}_{t-1}^T \overrightarrow{x}_i < 0. \end{cases} \tag{6}$$

(3) Return to step 2 as long as there is an incorrect classification, and repeat until all samples are correctly categorized.

This gradient descent iterative process is used by the perceptron. If the sample set can be separated linearly, there are

$$\overrightarrow{w}^T \overrightarrow{x} = \overrightarrow{w}_{t-1}^T \overrightarrow{x} + \eta \overrightarrow{x}^T \overrightarrow{x} > \overrightarrow{w}_{t-1}^T \overrightarrow{x}. \tag{7}$$

After a finite number of modifications, according to the perceptron criterion function in equation (2), we can prove that the algorithm must converge to a solution vector of $\overrightarrow{w}^*$. The perceptron algorithm, that is, the 1st that is executed by online learning in ML and can address the problem of linear separability, is a reward and punishment process. By directly extending the perceptron, we can get the second-order perceptron. The product of the perceptron and the multiplier of the current perceptron column are updated according to the product of the perceptron and the multiplier of the current perceptron column, and the weight of the former perceptron column is updated according to the product of the perceptron and the multiplier of the current perceptron column, and the weight of the former perceptron column is updated according to the product of the perceptron and the multiplier of the current perceptron column. Another second-order information online learning algorithm is the confidence weighting algorithm. This algorithm sets each feature with a different confidence. The weight of the feature with lower confidence is updated more aggressively when updating the weight. The weight of the feature is updated more conservatively if the feature has high confidence. The weight corresponding to each feature is generally assumed to be Gaussian distribution, so the second-order information is introduced into the model.

Since 2000, the convex model has become the mainstream of global optimization and verification. The online passive, active algorithm is a convex optimization-based online learning system. The online algorithm turns the constraint of the support vector machine's maximum interval into the constraint of identifying the closest neighbor

to the current classifier. When the distance between the new sample and the hyperplane of equation (3) is less than 1, the sample loss is generated, which is expressed by the hinge loss function of the following equation:

$$\ell\left(\overrightarrow{\mathrm{w}};\left(\overrightarrow{\mathrm{x}},y\right)\right)=\begin{cases}0, & y\left(\overrightarrow{w}^{T}\overrightarrow{x}\right)\geq 1,\\[2mm]1-y\left(\overrightarrow{w}^{T}\overrightarrow{x}\right), & \text{otherwise.}\end{cases} \quad (8)$$

In this way, the loss at t time of the sample is $\ell_{t}=\ell(\overrightarrow{w}_{t};(\overrightarrow{x}_{t},y_{t}))$. Suppose $\overrightarrow{w}_{t}=(0,\ldots,0)$; the learning weight vector is updated to solve the constrained optimization problem described in the following formula:

$$\overrightarrow{\mathrm{w}}_{t+1}=\operatorname{argmin}_{\overrightarrow{w}\in\mathbb{R}^{d+1}}\left(\frac{1}{2}\overrightarrow{w}-\overrightarrow{w}_{t}^{2}\right),s.t.\ell_{t}\left(\overrightarrow{w};\left(\overrightarrow{x}_{t},y_{t}\right)\right)=0. \quad (9)$$

It is not difficult to see that the online passive, active algorithm has two requirements: on the one hand, it requires $\overrightarrow{w}_{t+1}$ to correctly classify the current sample wood with a large enough spacing; on the other hand, $\overrightarrow{w}_{t+1}$ should be as close to $\overrightarrow{w}_{t}$ as possible. When $\ell_{t}=0$, then $\overrightarrow{w}_{t+1}=\overrightarrow{w}_{t}$. The algorithm derived from optimization equation (7) is "passive." On the contrary, if $\ell_{t}>0$, the algorithm "actively" urges $\overrightarrow{w}_{t+1}$ to meet the constraint with a loss of 0.

The Lagrangian convex optimization problem constructed by equation (9) is shown in the following equation:

$$\mathscr{L}\left(\overrightarrow{\mathrm{w}},\tau\right)=\frac{1}{2}\overrightarrow{w}-\overrightarrow{w}_{t}^{2}+\tau\left(1-y_{i}\left(\overrightarrow{w}^{T}\overrightarrow{x}_{i}\right)\right). \quad (10)$$

Here, $\tau\geq 0$ is the Lagrangian multipliers. See equation (11).

$$0=\nabla_{\overrightarrow{\mathrm{w}}}\mathscr{L}\left(\overrightarrow{w},\tau\right)=\overrightarrow{w}-\overrightarrow{w}_{t}-\tau y_{i}\overrightarrow{x}_{i}\Rightarrow\overrightarrow{w}=\overrightarrow{w}_{t}+\tau y_{i}\overrightarrow{x}_{i}. \quad (11)$$

Substitute equation (11) to equation (10), and we have the following result:

$$\mathscr{L}\left(\tau\right)=-\frac{1}{2}\tau^{2}\overrightarrow{x}_{i}^{2}+\tau\left(1-y_{i}\left(\overrightarrow{w}_{t}^{T}\overrightarrow{x}_{i}\right)\right). \quad (12)$$

Hence, we can get some results as follows:

$$\begin{aligned}0&=\frac{\partial\mathscr{L}\left(\tau\right)}{\partial\tau},\\&=-\tau\overrightarrow{x}_{i}^{2}+\left(1-y_{i}\left(\overrightarrow{w}_{t}^{T}\overrightarrow{x}_{i}\right)\right),\\\tau&=\frac{1-y_{i}\left(\overrightarrow{w}_{t}^{T}\overrightarrow{x}_{i}\right)}{\overrightarrow{x}_{i}^{2}},\\&=\frac{\ell_{t}}{\overrightarrow{x}_{i}^{2}}\left(\ell_{t}>0\right).\end{aligned} \quad (13)$$

According to the above analysis, the online passive, active algorithm of the binary classification problem can be obtained in Table 1. In short, the update rule of the online passive, active algorithm is that when the new data has no

TABLE 1: Online passive-active algorithm for classification problem.

Initialize: $w_1 = (0, \ldots, 0)$
For $t = 1, 2, 3, \ldots,$ do
Accept sample: $x_t \in \mathbb{R}^{d+1}$
Predict: $\widehat{y}_t = \operatorname{sgn}(w_t^T x_t)$
Receive the correct label: $y_t \in \{-1, +1\}$
Compute the loss: $l_t = \max\{0, 1 - y_t(w_t^T x_t)\}$
Update: $w_{t+1} = w_t + \tau_t y_t x_t, \tau_t = l_t/x_t^2$
End for

error, the algorithm updates passively; that is, it does not update; when the new data produces errors, the algorithm projects to the nearest neighbor of the existing classifier, that is, active update.

3. Parallel Community Partition Algorithm Based on MapReduce

The past decade has witnessed the extension of complex networks, an emerging research field. The Internet is everywhere, and as individuals, we are also a member of the social network. It can refer not only to the tangible network in real life, such as the subway system, but also to the definition in abstract space, such as the relationship network between friends. Over time, people's research focuses on the structural analysis of small networks to the analysis of systems with thousands or millions of nodes. The analysis of complex networks with massive data will face many great challenges. Presently, time complexity and division accuracy are two thorny problems in the division of human scale complex tree end communities. The community structure obtained by a fast algorithm is not accurate. The algorithm with accurate partition results often has high time complexity. Therefore, it is very important to find a fast and reliable social closure partition algorithm. MapReduce algorithm is favored. The MapReduce programming paradigm and algorithm include two major stages which are Map and Reduce.

The application of the community partition algorithm in this paper is a large-scale guarantee network. Some nodes can not be ignored in the analysis of the guarantee network. Which nodes need key prevention and control? Which nodes can be cleaned up temporarily? Therefore, this paper thinks of the scan algorithm based on structural similarity. The algorithm can not only identify the community structure but also find the cluster points and outliers in the community. Cluster points and outliers are the nodes that need to be monitored and can be ignored temporarily in the guarantee network. The SCAN algorithm, on the other hand, has a significant temporal complexity, especially for large-scale networks. The larger the number of edges of the processed tree network, the greater the time consumption. Because the idea of the SCAN algorithm is to traverse all nodes in the network and calculate the structural similarity between the parent node and its adjacent nodes, for the large-scale guarantee relationship data in this paper, the time consumption cost is very high. The calculation of structural similarity is to determine the core point. After calculation, the node pairs whose structural similarity does not meet the

TABLE 2: The algorithm for obtaining the adjacency table.

Input: relationship $\langle u, v \rangle$
Output: Node u and its ε-adjacency table $N_\varepsilon(u)$
Stage Map:
Input: Guarantee relationship $\langle u, v \rangle$
Output: $\{\langle u, v \rangle, \sigma(u, v)\}$
For each guarantee relationship, do
Key←u
Value←v
Output $\langle$key, value$\rangle$
Output $\langle$value, key$\rangle$
End for
Stage Reduce
Input: $\langle$key, value$\rangle$, key is the key$'$ in the Map1 stage, value is the value$'$ in Map1 stage
Output: $\langle$key$'$, value$'$ $\rangle$
For each $\langle$key, value$\rangle$ do
Key$'$←key
Value$'$←key ∪ value
Output $\langle$key$'$, value$\rangle'$
End for

conditions will not participate in the subsequent calculation. It can be seen that the larger the scale, the more the wasted time.

3.1. PFSCAN Algorithm Based on MapReduce.

The proposed improved algorithm is based on the scan algorithm. The scan algorithm deals with an undirected graph with no authority. Graph G = {V, E} is known, where V is the set of nodes in the graph and E is the set of node pairs, that is, edges in the graph. Considering the shortcomings of the scan algorithm in dividing large-scale networks, based on its idea and ensuring the accuracy of the division results, this paper proposes a parallel community division algorithm, PFSCAN. The algorithm carries out a pruning strategy, and parallel processing on the SCAN algorithm reduces the time of calculating the structural similarity of the original algorithm and makes full use of the advantages of MapReduce to further improve the partition efficiency.

Theorem 1. *Given node pair (v, w), if v, w satisfies $|\Gamma(v)| < (2\varepsilon - 1)|\Gamma(w)|$ or $|\Gamma(w)| < (2\varepsilon - 1)|\Gamma(v)|$, the structural similarity of (v, w) is similar to that of node u, v.*

According to Theorem 1, if $|\Gamma(v)| < (2\varepsilon - 1)|\Gamma(w)|$ or $|\Gamma(w)| < (2\varepsilon - 1)|\Gamma(v)|$ holds, it can be concluded that node pair (u, v) does not meet the condition of the ε-neighbor node. Therefore, before calculating the structural similarity, a judgment is added to "cut off" the node pairs that do not meet the conditions to reduce the amount of calculation of structural similarity. The main step of the pruning algorithm is to calculate the similarity of the remaining nodes. Then, if the structural similarity of the edges is less than the parameter, cut them off. Finally, discover the community, which is connected to the trimmed network. Cluster points or outliers are nodes that do not belong to any of the communities. In MapReduce, the three processes above can be run in simultaneously. We need to know the adjacency table of the two nodes to determine structural similarity. Because each edge's structural similarity computation is independent, it may be done in MapReduce at the same time.

Tables 2 and 3 are the description of the algorithms for obtaining the adjacency table and calculating the structural similarity. The MapReduce programming paradigm and algorithm include two major stages that are Map and Reduce. Tables 2 and 3 depicted the algorithm for obtaining the adjacency table and the algorithm for calculating the structural similarity. Both algorithms have the Map and Reduce stages. In the above algorithm with the given parameter E, the nodes whose structural similarity is greater than or equal are directly output.

The second step of the PFSCAN algorithm is realized. The last and most important step is to find the connected parts in the pruned network to get the final division result. Before execution, process the results of the previous step and output the results of each node ε-neighbor node, that is, the neighbor node whose structural similarity meets the conditions. The specific algorithm description is shown in Table 4.

Step three can be implemented in parallel with the help of the idea of a label propagation algorithm. The parallelism is achieved with the MapReduce programming paradigm integrated with the Hadoop platform. The basic idea is to define a label for each node during initialization. Through the "propagation" of the label, the nodes with the same label belong to the same community. To realize label propagation, first define a piece of node information for each node, including status, label, and ε, a neighbor node. Each node has two states, active and inactive. The active node needs to propagate its label to its neighbor nodes, while the inactive node does not need to do so. During initialization, each node is defined as the active state, and their node number is defined as the initial label. The initialization algorithm is shown in Table 5.

TABLE 3: The algorithm for calculating the structural similarity.

Input: node u and its adjacency table $N(u)$, parameter ε

Output: The relationship $\langle u, v \rangle$ whose structure similarity is greater than ε and its structure similarity $\sigma(u, v)$

Map Stage:

  Input: $\langle \mathbf{key}, \mathbf{value} \rangle$, key is the offset of the document, valuestands for $\{\langle u, v \rangle, \sigma(u, v)\}$

  Output: $\langle \mathbf{key}', \mathbf{value}' \rangle$

  For each value, do

    For each $v \in N(u)$ do

      If $v < u$ then

        Key$' \leftarrow \langle v, u \rangle$

      Else

        Key$' \leftarrow \langle u, v \rangle$

      End if

    End for

    Value$' \leftarrow u$: $N(u)$

    Output $\langle$key$'$, value$'\rangle$

  End for

Reduce Stage:

  Input: $\langle \mathbf{key}, \mathbf{value} \rangle$, key is the key$'$ in the Map stage, value is the value$'$ in the Map stage, parameter ε

  Output: $\langle \mathbf{key}', \mathbf{value}' \rangle$

  For each $\langle \mathbf{key}, \mathbf{value} \rangle$ do

    Key$' \leftarrow$ key

    If the value contains only one adjacency table, then

      Value$' \leftarrow 1$

      Output $\langle$key$'$, value$'\rangle$

    Else

      Obtain the two adjacency table $N(u)$ and $N(v)$ contained in the value

      If $|N(u)| < (2\varepsilon - 1)|N(v)|$ then

        Return

      Else

        $\sigma(u, v) = |N(u) \cap N(v)| / \sqrt{|N(u)||N(v)|}$

        If $\sigma(u, v) > \varepsilon$ then

          Value$' \leftarrow \sigma(u, v)$

          Output $\langle$key$'$, value$'\rangle$

        Else

          Return

        End if

      End if

    End if

  End for

---

4. Experiments and Results

This section will carry out experiments from three aspects to verify the accuracy, efficiency, and parallel performance of PFSCAN. The prosed model is implemented using the Apache Hadoop platform. The Hadoop platform is a parallel and distributed platform that is used to analyze and process big data. The rationale of Hadoop is the parallel processing. The MapReduce programming paradigm is used for the programming. MapReduce is a functional programming paradigm. It is based on parallel programming concepts. The experiment's algorithms were written in the Java programming language. The Java Mapper and Reducers classes are used for the programming. Table 6 depicts the specific experimental environment.

Now, the three general approaches utilized to measure the results of community division are modularity Q, standardized mutual information NMI, and adjusted Rand coefficient ARI. The prosed model integrates the modularity Q and the standardized mutual information NMI. Block degree Q was first defined by Newman. The block degree Q method is used to evaluate the partition level of unknown networks. Modularity is the probability expectation obtained by subtracting the proportion of edges belonging to the same community from the random distribution of these edges. The specific definition is as follows:

$$Q = \frac{1}{2m} \sum_{vm} \left[A_{vm} - \frac{k_v k_m}{2m} \right] \delta_{vm}. \tag{14}$$

The NMI is used to compare the real-world community structure to the community structure generated by the clustering technique. In other words, it is best for networks with well-established community structures. Equation (15) is written as follows:

$$NMI = \frac{\left(-2 \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} C_{ij} \log\left(C_{ij} N / C_i C_j \right) \right)}{\sum_{i=1}^{C_A} C_i \log\left(C_i / N \right) + \sum_{j=1}^{C_A} C_i \log\left(C_i / N \right)}. \tag{15}$$

TABLE 4: The EPS_AdjacencyList algorithm.

Input: Relationship $\langle u, v \rangle$ and their structure similarity $\sigma(u, v)$
Output: Node u and its ε-adjacency table $N_\varepsilon(u)$
Stage Map:
Input: $\langle$key, value$\rangle$, key is the offset of the document, valuestands for$\{\langle u, v \rangle, \sigma(u, v)\}$
Output: $\{\langle u, v \rangle, \langle v, u \rangle\}$
For each value, do
Key$' \leftarrow u$
Value$' \leftarrow v$
If $u = v$ then
Output $\langle$key$'$, value$'\rangle$
Else
Output $\langle$key$'$, value$'\rangle$
Output $\langle$value, key$\rangle$
End if
End for
Stage Reduce:
Input:$\langle$key, value$\rangle$,key is the key$'$ in the Map1 stage, valueis thevalue$'$ in Map1 stage
Output:$\langle$key$'$, value$'\rangle$
For each $\langle$key, value$\rangle$ do
Key$' \leftarrow$key
Value$' \leftarrow$value
Output $\langle$key$'$, value$'\rangle$
End for

TABLE 5: The node information algorithm.

Input: node u, ε-adjacency table, parameter μ
Output: Node u and Node information information (status, label, $N_\varepsilon(u)$)
Map1 Stage:
Input: $\langle$key, value$\rangle$, keyis the offset of the document, valuestands for $\{u, N\_\varepsilon(u)\}$
Output: $\{u, N_\varepsilon(u),$ null/nonmember$\}$
For each value, do
Key$' \leftarrow u$
If $|N_\varepsilon(u)| \geq \mu$ then
Value$\prime \leftarrow$information $N_\varepsilon(u)$
Else
Value$' \leftarrow$information. $N_\varepsilon(u) \cup$ nonmember
End if
End for
Map2 stage:
Input: $\langle$key, value$\rangle$, keyis the key$'$in the Map1 stage, valueis thevalue$'$ in Map1 stage
Output: $\langle$key$'$, value$'\rangle$
For each $\langle$key, value$\rangle$ do
If value contains nonmember then
Information label$\leftarrow\ = -1$
Else
Information label$\leftarrow$key
End if
Information.status$\leftarrow$activated
Key$\leftarrow$key
Value$\prime \leftarrow$information.status $\cup$ information.label $\cup$ information.$N\_\varepsilon(u)$
Output $\langle$key$'$, value$'\rangle$
End for

To verify the accuracy of the PFSCAN algorithm, this paper first cites two classical real data sets in the algorithm. In the experiment of community discovery algorithm, elegant data of neural network and dolphin network are used as the test network. The worm nematode neural network contains 297 nodes and 2345 arcs; the Dolphin network is a

TABLE 6: Experimental environment.

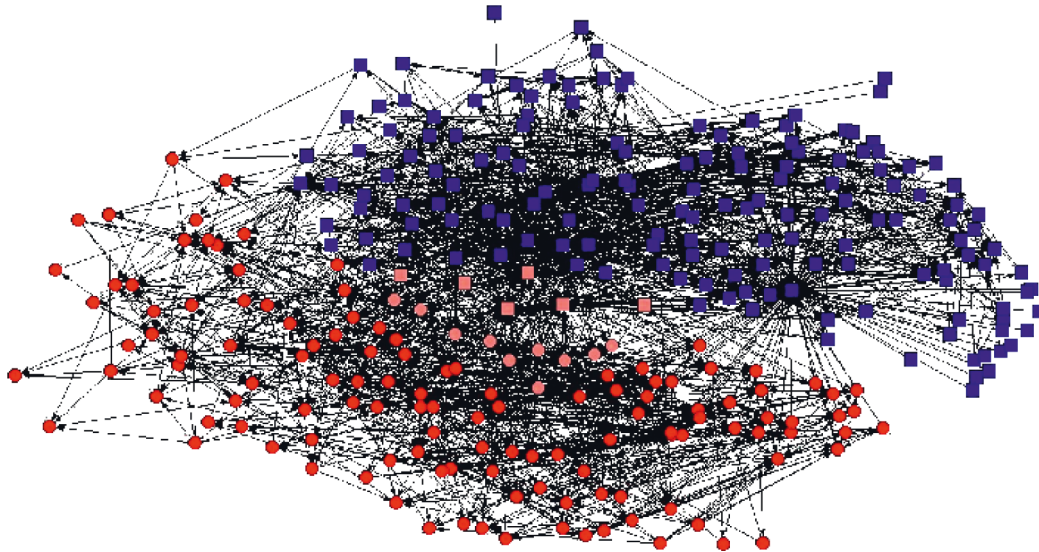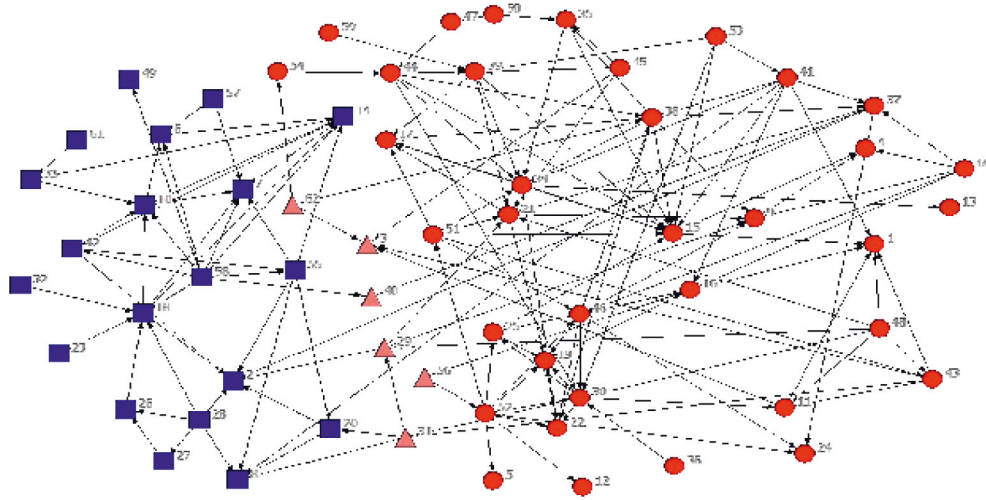| Environment	Configuration
Processor	Inter (R) Xeon (R) Silver 4114
Memory	64 GB
MapReduce framework	Hadoop 2.7.1
Operation system	Ubuntu 16.04 LTS
Number of clusters	4

FIGURE 1: Findings of worm nematode neural network community.

FIGURE 2: Dolphin network community discovery results.

TABLE 7: Large-scale network datasets.

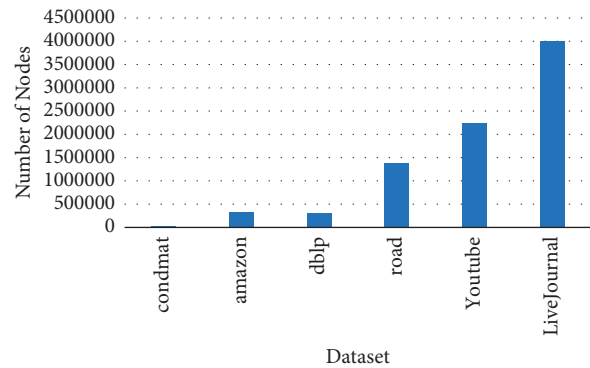| Dataset	Nodes	Links	Graph type	C
Condmat	23133	186936	Social network	0.6334
Amazon	334863	925872	Web graph	0.3967
Dblp	317080	1049866	Social network	0.6324
Road	1379971	3843320	Road network	0.0470
Youtube	2244789	5894561	Social network	0.5143
Livejournal	4007568	578945125	Social network	0.2846

FIGURE 3: Datasets nodes.

classic data set in social network analysis. It describes the "social relationship" formed between 62 dolphins living in New Zealand. Each dolphin is represented by a node in the network. Dolphins with frequent communication with each other will produce an edge. Therefore, the 62 nodes of the dolphin network produce 159 edges. The clustering results are shown in Figures 1 and 2.

The partition effect of the technique described in this research is perfectly consistent with the SCAN algorithm, as shown by the clustering findings of the aforementioned two data sets. It is critical to test the operating efficiency of this
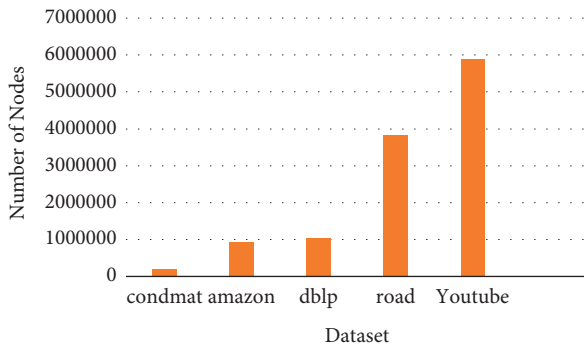
FIGURE 4: Datasets links.

algorithm after checking its accuracy. PFSCAN is used to partition certain large-scale network data sets to verify the advantages of this technique in dealing with large-scale data. The running time of the division result is compared with the SCAN algorithm and improved algorithm LINKSCAN*. These large-scale network data sets are from standard, and their structural characteristics are shown in Table 7. In addition, the dataset nodes and links information is also depicted in Figures 3 and 4.

5. Conclusion

This article reviews the network structure of the guarantee network and finds the core guarantee group, cluster point, and outlier point in the network using the complex network theory. A dynamic model suitable for the guarantee network is constructed based on the user model on the connected guarantee subnetwork. The model is simulated, and the effects of different parameters and infectious sources on risk propagation in the guarantee network are discussed to formulate corresponding crisis supervision measures and contribute to reducing risk costs. China is currently in a strategic window of opportunity to deepen reform and achieve economic and social transformation. The contrast between rapid economic progress and the relative lag in social construction, on the other hand, has become increasingly apparent. In the context of big data, promoting mechanism development and technical innovation at the same time can compensate for any flaws created by institutional ambiguity and provide a stronger assurance for social governance refinement. The proposed model is established for the big data thinking paradigm to promote social risk governance innovation. The proposed model also realizes diversified and coordinated risk governance. The results reveal the significance of the proposed model.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that he has no conflicts of interest.

References

[1] L. M. S Ni, X. Xiao, and H. Tan, "The golden age for popularizing big data," *Science China Information Sciences*, vol. 59, no. 10, Article ID 108101, 2016.

[2] Z. Liu, Z. Deng, S. J. Davis, C. Giron, and P. Ciais, "Monitoring Global Carbon Emissions in 2021," *Nature Reviews Earth & Environment*, vol. 3, no. 4, pp. 1–3, 2022.

[3] Y.. Zhu, "New national initiatives of modernizing education in China," *ECNU Review of Education*, vol. 2, no. 3, pp. 353–362, 2019.

[4] M. Chen, S. Mao, and Y. Liu, "Big data: a survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.

[5] N. Khan, I. Yaqoob, H Ibrahim Abaker Targio et al., "Big data: survey, technologies, opportunities, and challenges," *The Scientific World Journal*, vol. 2014, 18 pages, 2014.

[6] X. Li, H. Liu, W. Wang, Y. Zheng, H. Lv, and Z. Lv, "Big data analysis of the internet of things in the digital twins of smart city based on deep learning," *Future Generation Computer Systems*, vol. 128, pp. 167–177, 2022.

[7] M. Naeem, T. Jamal, J. Diaz-Martinez et al., "Trends and future perspective challenges in big data," *Advances in Intelligent Data Analysis and Applications*, pp. 309–325, 2022.

[8] S. Rogers Van Katwyk, S. L. Jones, and S. J. Hoffman, "Mapping educational opportunities for healthcare workers on antimicrobial resistance and stewardship around the world," *Human Resources for Health*, vol. 16, no. 1, p. 9, 2018.

[9] S. Dodge, "A data science framework for movement," *Geographical Analysis*, vol. 53, no. 1, pp. 92–112, 2021.

[10] L. Tian, "What is big data," *Data Science Central*, vol. 5, 2012, https://www.ibm.com/big-data/us/en/.13DiyaSoubra. The3vsthatdefinebigdata.

[11] S. R. M. Zeebaree, H. M. Shukur, L. M. Haji, R. R. Zebari, K. Jacksi, and S. M. Abas, "Characteristics and analysis of hadoop distributed systems," *Technology Reports of Kansai University*, vol. 62, no. 4, pp. 1555–1564, 2020.

[12] H.-C. Lu, F. J. Hwang, and Y.-H. Huang, "Parallel and distributed architecture of genetic algorithm on Apache Hadoop and Spark," *Applied Soft Computing*, vol. 95, Article ID 106497, 2020.

[13] A. Mostafaeipour, A. Jahangard Rafsanjani, M. Ahmadi, and J. Arockia Dhanraj, "Investigating the performance of Hadoop and Spark platforms on machine learning algorithms," *The Journal of Supercomputing*, vol. 77, no. 2, pp. 1273–1300, 2021.

[14] W. Xu, X. Zhao, B. Lao, and G. Nong, "Enhancing HDFS with a full-text search system for massive small files," *The Journal of Supercomputing*, vol. 77, no. 7, pp. 7149–7170, 2021.

[15] V. S. Sharma and N. C. Barwar, "Performance evaluation of merging techniques for handling small size files in HDFS," *Data Analytics and Management*, pp. 137–150, 2021.

[16] R. Abdolazimi, M. Heidari, A. Esmaeilzadeh, and N. Hassan, "Mapreduce preprocess of big graphs for rapid connected components detection," in *Proeedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0112–0118, IEEE, Manhattan, NY.USA, 2022.

[17] F. Ma, F. Lv, P. Xu et al., "Task shifting of HIV/AIDS case management to Community Health Service Centers in urban China: a qualitative policy analysis," *BMC Health Services Research*, vol. 15, no. 1, p. 253, 2015.

[18] F. N. Pieke, "The Communist Party and social management in China," *Critical Readings on the Communist Party of China (4 Vols. Set)*, vol. 4, pp. 998–1018, 2017.

[19] L. Zhao and Y. Li, "Crowdfunding in China: turmoil of global leadership," *Advances in crowdfunding*, vol. 12, pp. 273–296, 2020.

[20] J. Li, C. Gu, Z. Wu, and T. Huang, "Online learning algorithm for distributed convex optimization with time-varying coupled constraints and bandit feedback," *IEEE Transactions on Cybernetics*, vol. 52, no. 2, pp. 1009–1020, 2022.

[21] H. Zhao and C. Zhang, "An online-learning-based evolutionary many-objective algorithm," *Information Sciences*, vol. 509, pp. 1–21, 2020.

[22] Z. Sun and M. R. Nakhai, "An online learning algorithm for distributed task offloading in multi-access edge computing," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3090–3102, 2020.