Hindawi
Security and Communication Networks
Volume 2022, Article ID 3826126, 9 pages
https://doi.org/10.1155/2022/3826126

Research Article

Research on Privacy Protection Technology of Mobile Social Network Based on Data Mining under Big Data

Jiawen Du [1] and Yong Pi [2]

[1] *Law School, Wuhan University, Wuhan 430072, Hubei Province, China*
[2] *Shanghai International College of Intellectual Property, Tongji University, Shanghai 200092, China*

Correspondence should be addressed to Jiawen Du; djwkira@whu.edu.cn

Received 24 November 2021; Accepted 18 December 2021; Published 13 January 2022

Academic Editor: Thippa Reddy G

Copyright © 2022 Jiawen Du and Yong Pi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the advent of the era of big data, people's lives have undergone earth-shaking changes, not only getting rid of the cumbersome traditional data collection but also collecting and sorting information directly from people's footprints on social networks. This paper explores and analyzes the privacy issues in current social networks and puts forward the protection strategies of users' privacy data based on data mining algorithms so as to truly ensure that users' privacy in social networks will not be illegally infringed in the era of big data. The data mining algorithm proposed in this paper can protect the user's identity from being identified and the user's private information from being leaked. Using differential privacy protection methods in social networks can effectively protect users' privacy information in data publishing and data mining. Therefore, it is of great significance to study data publishing, data mining methods based on differential privacy protection, and their application in social networks.

1. Introduction

In recent years, with the development of communication technology, social software has brought convenience to user communication, ensured real-time contact between users, and accelerated the dissemination of information and news. Therefore, more and more users are attracted to register and use them. Their social circles have moved to social platforms, and their various activities and behaviors on social platforms have accumulated a lot of data [1]. With the rapid development of database technology and the continuous improvement of hardware level, as well as the increasing demand for information dissemination and sharing, a large amount of useful data can be saved [2]. Faced with such massive data storage, data mining and data publishing have become two important research directions for database applications [3]. Data mining is intended to extract meaningful rules and models from data, and data publishing is to present the data in an appropriate form [4]. Data release and data mining in social networks are likely to cause the personal sensitive information included in the social network and the relationship between users to be destroyed or information leaked, which greatly affects the security of the use of social networks. There is a great risk of privacy leakage [5]. Therefore, how to better publish and mine the massive information in social networks without destroying its private information has become an important research topic in social networks [6].

Continuously enhancing the security of social networks and continuously improving privacy protection capabilities will help people use social networks more safely and securely [7]. At present, many privacy protection technologies have been proposed for user privacy and security issues in social networks. The easiest way to implement the technology is only to hide user identity information and not to process other information [8]. Although this technology protects the user's personal privacy within a certain range, malicious actors can still identify the individual's identity through the background knowledge of the target user's social network relationship, leading to the disclosure of user privacy [9]. Therefore, how to ensure the privacy and security of users when performing data mining on social network data is of great significance [10]. The social network recommendation system not only helps users find valuable information for themselves but also allows the information to be displayed to interested users so as to achieve a win-win situation for information producers and information consumers [11]. Privacy protection mainly includes two aspects: the protection of sensitive knowledge and the protection of sensitive data. Sensitive knowledge mainly refers to sensitive knowledge such as association rules and classification rules extracted from the database; sensitive data refers to the private data that can correspond to an individual, thereby causing the individual to be exposed [12]. The paper explores and analyzes the privacy issues in current social networks and proposes user privacy data protection strategies based on data mining algorithms so as to hope that in the era of big data, the privacy of users in social networks will no longer be illegally violated [13].

Mobile social networking has become a rapidly growing application among domestic and foreign mobile users. It is urgent to protect user privacy [14]. The existing simple data processing methods cannot meet the needs of privacy protection, and the existing laws and regulations have restricted the application and development of data mining technology [15]. If certain protection measures are not taken for the information, the private information of a specific individual will be exposed, which will cause harm to the owner of the data. Similarly, if the protection measures taken are improper or too simple, then reasonable data mining methods will be used to obtain the private information of a specific individual, resulting in privacy leakage [16]. The privacy protection of social network data is to perform some artificial operations on the original network data, such as adding, deleting, or modifying parts, so that the attacker cannot obtain the user's sensitive information and avoid information leakage [17]. The data mining algorithm proposed in this paper can well protect the user's identity from being identified and the user's private information from being leaked. The algorithm can decompose the data, reconstruct the features, and store the data vertically, which can effectively prevent the data from being threatened by security and will not cause the loss of mining accuracy. Only the processed data can be released to the public. Of course, while protecting the user's sensitive information, making the processed information still have certain usability is also an important factor in measuring data anonymity.

2. Related Work

Literature [18] classifies privacy protection technologies into three categories according to different specific applications. They are privacy protection based on data perturbation technology, data encryption, and data anonymization.

Literature [19] put forward the concept of database anonymization and used the generalization method to hide sensitive attributes in groups of scale.

Literature [20] proposed a k-degree model for privacy protection of node degrees in social networks, which made it impossible for attackers to identify target nodes by collecting node degrees as background knowledge.

Literature [21] proposed to minimize information loss while generating a k-degree anonymous model.

Literature [22] proposed to construct the k-degree anonymous graph by using the idea of dynamic programming to protect the privacy of social network structure.

Literature [23] proposed that the parameter k of many existing K-anonymity models is predefined, and K represents the privacy protection of nodes in social networks. The idea of personalized privacy protection is formally introduced, and a K-anonymity model based on personalized privacy protection requirements is proposed.

Literature [24] divided the original network into k isomorphic subgraphs, which effectively prevented the node reidentification attack.

Literature [25] constructs K anonymity model for path privacy, and the construction method is to modify different types of edges based on greedy ideas.

Literature [26] combines L diversity on the basis of the k-degree model to protect the sensitive attributes of nodes or edge relations in social networks.

In literature [27], through clustering technology, the nodes in the original network are clustered to obtain an anonymous network composed of super nodes, and the super nodes are generalized to achieve the purpose of privacy protection.

Due to the development of technology sharing, big data are widely used in every aspect of life, and unreasonable use also brings great troubles and even terrible threats to users. However, at present, there is still no mature technology and relevant perfect laws and regulations for the protection of users' privacy. The lack of this aspect makes it impossible for relevant industry standards to have clear boundaries and implementation criteria and to implement effective measures to overcome this shortcoming. In order to ensure users' privacy, this paper carries out effective data mining and analysis on social networks. Combined with the KD tree optimization algorithm, a social network model based on data mining is built to protect the privacy of social networks, and experimental verification and algorithm analysis are carried out on data sets.

3. Methodology

Big data is like a huge spider web, weaving the network information of today's society. It is a large-scale and quite complex project, with the collection and processing function irreplaceable by other modern technologies. Thus, complexity, diversity, scale, and convenience are the outstanding characteristics of big data. It is such a combination of characteristics that big data technology has incomparable advantages over other technologies. The main problem of attribute and relationship-oriented data privacy protection is how to hide data in a relational database. The three common directions are data anonymization, secure multiparty computing, and data distortion. The comprehensive application of the three directions can effectively reduce the risk of personal data leakage.

The goal of the anonymous triangle protection principle is to protect those anonymous triangles in the process of

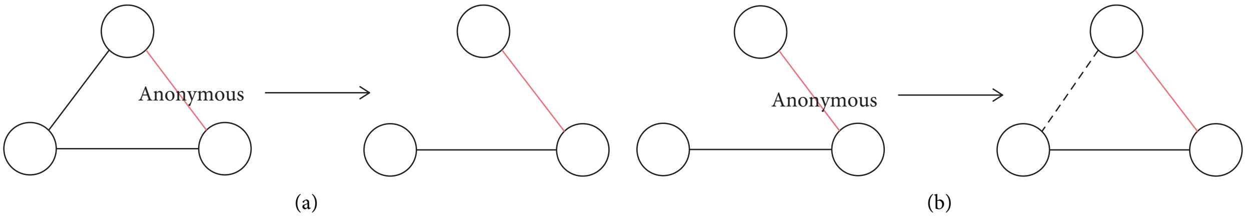

Figure 1: Anonymization triangle protection principle: (a) delete edge and (b) add edge.

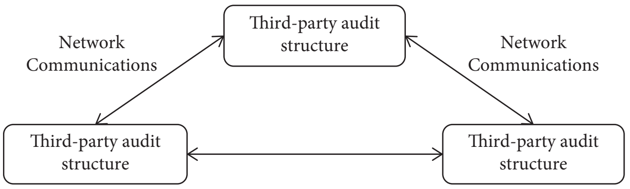
Third-party audit structure
Network Communications
Network Communications
Third-party audit structure
Third-party audit structure

Figure 2: Data storage security architecture based on cloud computing.

graph anonymization. If multiple edges are generated, the original triangle that has been anonymized will not be included, and the triangle will not be anonymized gradually, so as to protect the original relationship as shown in Figure 1.

The common friends in the social network are all distributed on a scale, so only a small part of the connected edges have a higher relationship value. In order to participate in convenient social activities and enjoy all-round customized services in the era of big data, users cannot have absolute privacy rights [28]. But this does not mean that social network users can relax the protection of personal privacy but should pay more attention to the awareness of personal privacy protection. Only in this way can we ensure that we can enjoy normal services and social activities in the torrent of the big data era and can protect our privacy from being violated. At present, the release of dynamic social network data divides the privacy protection needs in social networks into different levels and at the same time provides privacy protection for users' sensitive attributes and sensitive edges in social networks. The data storage security system architecture based on cloud computing is shown in Figure 2.

K-anonymity technology has been widely used in anonymous relational data. In the privacy protection of graph data, many researchers still use *k*-anonymity technology to expand its application to graph data. *K*-nearest neighbor anonymity extracts all nodes with similar neighbors, encodes them, and divides them into the same group until each group is composed of at least *k* nodes. Then each group is anonymized so that any node in the same group has at least $k-1$ isomorphic neighbor nodes. This method can effectively resist neighborhood attacks. For social networks, social networks have the characteristics of a "small world," and nodes with the same background are more likely to generate connections and aggregate in a small group. Therefore, the anonymous data after clustering privacy protection still retains the macro characteristics of the original network. Social network analysts can carry out data mining on social networks on the premise of ensuring users' privacy and security so as to ensure the effectiveness of anonymous data.

Traditional data mining refers to the process of discovering new knowledge based on the original data and using corresponding mining algorithms. Traditional data mining algorithms cannot effectively protect private data, and security is affected. The KD3 framework is based on traditional data mining technology to process the privacy data that needs to be protected to form a new release database D′. Then reconstruct its features to form a new data feature F. And use the algorithm on it to adjust to get a new data mining algorithm M′. Finally, get a new mining result X′, make X′ and X as close as possible. In this way, privacy data is effectively protected, and almost consistent mining results are obtained. The frame is shown in Figure 3.

At present, the development and utilization of various social software include the signing of privacy treaties, but most of these treaties are mandatory terms, and users can only be forced to accept them. Social network users cannot check individual options in the privacy terms according to their actual situation, so they check "agree" in order to have the right to use the software. In the process of software development, each merchant should take more initiative to consider the initiative of user authorization, rather than blindly forcing users to accept terms. Clustering-based privacy protection is also one of the mainstream protection technologies of graph data. The idea of aggregation is to aggregate the points or edges in the social network into a super point or super edge according to the similarity and perform the same anonymous operation on the members in the super point or super edge. Figure 4 shows the structure of the intrusion detection system.

Describe computer intrusion data by ω and ν. Among them, ω represents the horizontal domain vector of computer network intrusion data, and ν represents the vertical domain vector of computer network intrusion data. α represents the initial filtering result of the intrusion feature data; then α is expressed as follows:

$$\alpha = \sqrt{W}\omega \cdot s(\omega) + s(\nu) + m, \tag{1}$$

where W represents the norm vector of the intrusion signal, $s(\omega)$ represents the norm coefficient of the horizontal domain vector, $s(\nu)$ represents the norm coefficient of the vertical domain vector, and m represents the initial filtering constant. The signal processing result of the intrusion feature data can be expressed as follows:

$$R = W + 2\alpha\varsigma(n) \cdot \omega\nu, \tag{2}$$

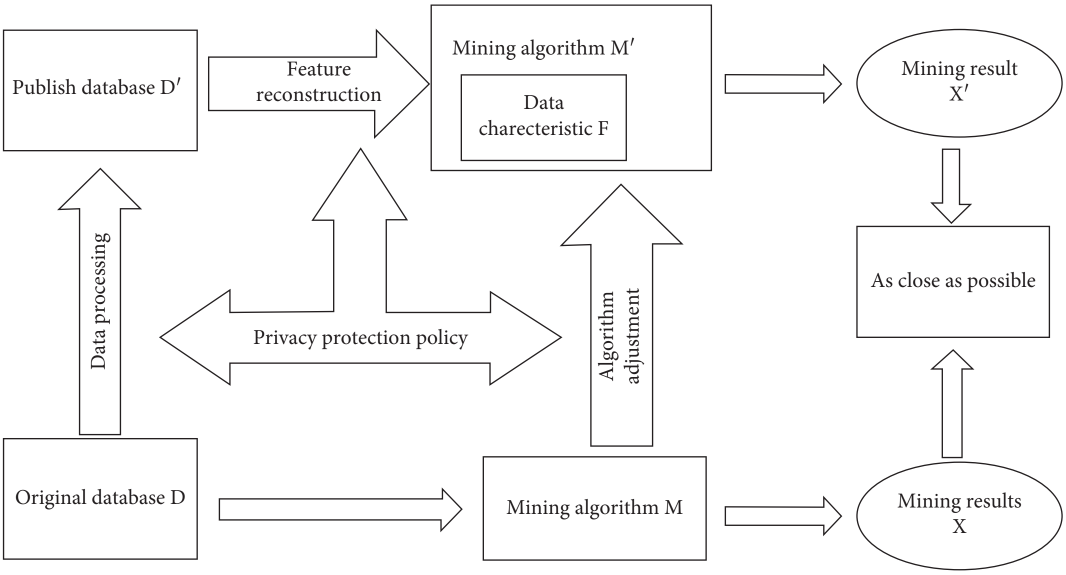

Figure 3: Data mining method framework for privacy protection.

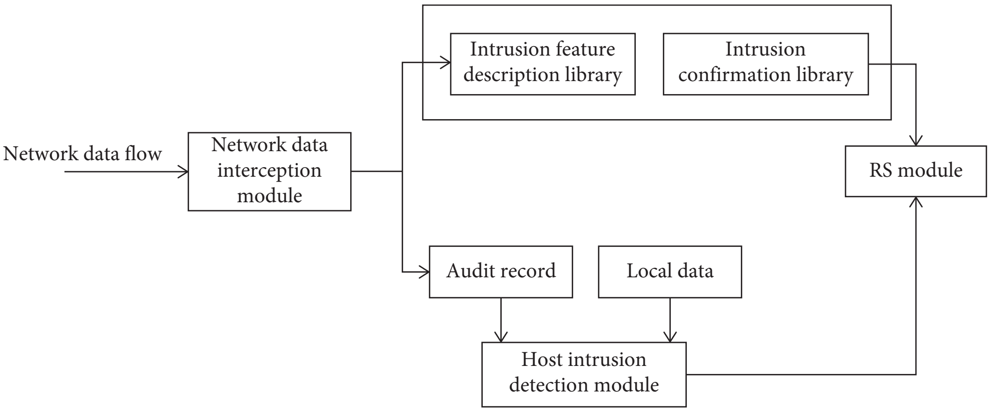
Network data flow
Network data interception module
Intrusion feature description library
Intrusion confirmation library
RS module
Audit record
Local data
Host intrusion detection module

Figure 4: Schematic diagram of the structure of the intrusion detection system.

where $\varsigma(n)$ represents the superimposed signal processing result of computer intrusion communication data.

The agglomeration coefficient is generally used in social networks to describe the degree of interconnection between a point and its neighboring points, that is to say, the agglomeration coefficient can reflect the degree of mutual understanding between a user's friends. The local agglomeration coefficient is used to describe the properties of a specific vertex, and the average agglomeration coefficient is used to describe the average of the local agglomeration coefficients of all vertices in the entire social network. In the social network $G = (V, E, \mathrm{L})$, G is an undirected graph. The local agglomeration coefficient C_i of a vertex V_i in G is shown in the following formula:

$$C_i = \frac{2\left|e_{jk}\right|}{k_i\left(k_i - 1\right)} v_j, \quad v_k \in N_i,\ e_{jk} \in E, \tag{3}$$

where e_{jk} is the edge between vertices i and j, and N_i is $N_i = \left\{v_j\colon e_{ij} \in E \cap e_{ji} \in E\right\}$, which is the set of adjacent vertices of vertex v_i. k_i is the number of adjacent vertices in v_i; then in the social network $G = (V, E, L)$, the average agglomeration coefficient is shown in the following formula:

$$CC = \frac{1}{n} \sum_{i=1}^{n} C(i), \tag{4}$$

where n is the number of vertices in social network G.

Social network researchers can still use the clustered graph features to investigate the macro characteristics of the original graph. The main idea of the algorithm is: cluster the nodes of the social network according to the comprehensive distance between the nodes, cluster them into several super points, and the specific details in the super points are hidden. As long as the nodes in the two super points have one edge connected, there is only one edge connected between the two super points.

In the social network $G = (V, E, L)$, the average path length APL is the average of the shortest distance between all vertices, as shown in the formula:

$$\mathrm{APL} = \frac{2}{n(n-1)} \sum_{v_i, v_j \in G} d\left(v_i, v_j\right), \tag{5}$$

where $d(v_i, v_j)$ is the shortest distance between the vertices v_i and v_j and n is the number of vertices in the social network G.

With the rapid development of the Internet and information technology, all kinds of data in social networks are constantly accumulating. With the progress of the times and the passage of time, big data has spread all over various fields and platforms. It also ushered in the generation of massive data. In the social network-oriented application, it is particularly important to protect users' privacy. By adopting certain protection strategies, users' data cannot be leaked, and their security can be guaranteed. KD tree is a kind of data structure, which can be used to divide data nodes into K-dimensional space. KD tree is a binary tree in which each node represents a spatial range. In order to further study the KD tree optimization center point selection method, define the following formulas. First, set the sample data set $A\{a_1, a_2, \ldots a_n\}$.

The number of data elements contained in a single rectangular cell Num

$$\mathrm{Num} = \frac{n}{m \times k}, \tag{6}$$

where n represents the number of elements in the sample data set, k represents the number of clusters, and m represents the number of sub-blocks contained in a cluster. The data can be adjusted timely according to the size of the given data set. Usually, when there is little difference in the number of data set samples, m can be taken as 10. A complete KD tree can be constructed by knowing the three parameters of N, M, and K, while the parameters of K and M represent the depth of KD tree and the number of contained leaf nodes, respectively.

Rectangular unit center C_i

$$c_i = \frac{S_i}{W_i}, \tag{7}$$

where S_i represents the linear sum of all elements in the rectangular unit, W_i represents the weight of the rectangular unit, and its value mainly represents the number of sample elements contained in the rectangular unit.

The density Den of the rectangular unit is mainly used to indicate the density between the data elements contained in the rectangular unit.

$$\begin{aligned}\mathrm{Den}_i &= \frac{W_i}{V_i} \\ &= \frac{W_i}{\left(\max\left(d_{\max} - d_{\min}\right)\right)^2},\end{aligned} \tag{8}$$

where W_i represents the number of sample elements contained in the rectangular cell, V_i represents the area of the rectangular cell, and $d_{\max}$ and $d_{\min}$ represent the maximum and minimum data elements in the corresponding rectangular cell, respectively.

With the improvement of data sharing and the development of data mining technology, people are getting more information, and the leakage of personal privacy data is getting more and more attention. The hierarchical information security organization is shown in Figure 5.

The existence of vertices is one of the most basic privacy information in social networks. Everyone may be on many different social networks. The same user may disclose different privacy in different social networks. Vertex is a necessary condition for the existence of a social network, and the attribute of the vertex is easy to obtain information in the social network graph. Although differential privacy protection can effectively protect users' social relations, it is mainly based on that the attacker has mastered some information about the attack object. Therefore, the ability of the attacker should be reasonably evaluated before designing the privacy protection algorithm.

4. Result Analysis and Discussion

Data mining is a process of extracting hidden patterns from data. It is an important way to transform data into information and knowledge, and it is one of the effective means to analyze and process large amounts of data. At present, data mining technology has been widely used in biology, natural language processing, information retrieval, and other fields. Applying data mining methods to the research of social networks has become a new branch in the field of data mining.

The above mainly introduces some basic theories in social networks and the background knowledge that an attacker may have to launch an attack. And the algorithm of KD tree optimization to select the center point is analyzed experimentally. Brief introduction and summary of structured privacy protection technology and privacy protection technology with label attribute data. Although there are endless methods to protect user privacy in social networks, with the vigorous development of social applications and the large-scale increase in the number of people using social networks, social network data will become more and more complex, and privacy protection technologies need to be more perfect. The efficiency and usability of the algorithm for selecting the initial center point based on the KD tree optimization are analyzed. All the experimental results are simulated in MATLAB. The data set used in the experiment comes from UCI Machine Learning Repository, and the five data sets used in UCI are Iris, Ecoli, AcuteInflammations, Breastcancer, and Thyroid for related research. Table 1 is a description of these five data sets.

The accuracy of the KD tree optimization center point selection algorithm and the traditional K-medoids clustering algorithm when performing the same clustering are compared, as shown in Table 2.

From Table 2, we can see that compared with the traditional K-medoids algorithm, the accuracy of the KD-tree optimized center point selection algorithm proposed in this paper has been significantly improved, which shows that the KD-tree optimized center point selection algorithm is very effective. However, in the experiment, due to the KD tree optimization algorithm, it is necessary to build a KD tree and calculate the center and density of rectangular elements, so the time consumption is relatively large, which is inevitable.

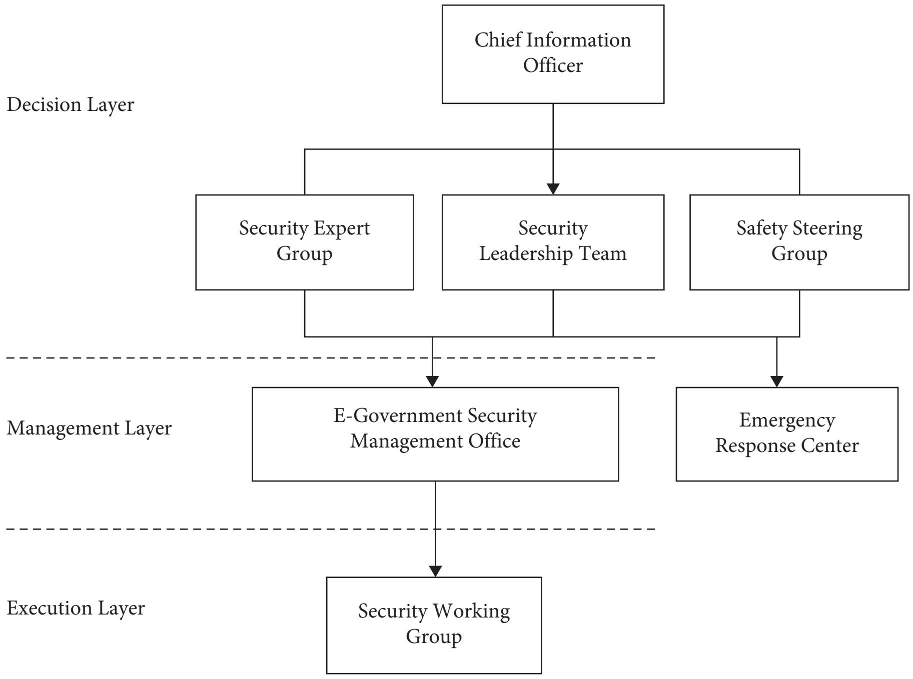

Figure 5: Hierarchical information security organization.

Table 1: Composition description of data set.

Data set name	Type of data	Number of records in the data set	Number of attributes	Number of clusters
Iris	Multivariate	152	5	4
Ecoli	Multivariate	334	7	8
Breastcancer	Multivariate	697	9	3
Thyroid	Multivariate	222	7	5

Table 2: Accuracy analysis of experimental results of KD tree optimization selection algorithm and traditional *K*-medoids algorithm.

	K-medoids algorithm		KD tree optimization algorithm	
	Running time (ms)	Accuracy (%)	Running time (ms)	Accuracy (%)
Iris	36	76.42	45	87.12
Ecoli	79	73.66	108	85.54
Breastcancer	84	93.24	99	96.21
Thyroid	74	79.36	86	84.53

Therefore, the KD tree optimization algorithm proposed in this paper has a high accuracy for the processing of data with low dimensions.

Next, we further verify the effectiveness of the algorithm for higher dimension data. Table 3 shows the attribute description of related data sets.

Table 3: Data set description.

Number of data sets	Data dimension	Number of clusters
$D1$	10	14
$D2$	20	14
$D3$	30	14
$D4$	40	14
$D5$	50	14

The above data are applied to the KD tree optimization selection algorithm proposed in this paper and the traditional *K*-medoids algorithm, and five independent experiments are performed on each group of data to analyze the accuracy rate in detail and select each group of data. The average of the results of the five experiments was recorded, and the results of the experimental analysis are shown in Figure 6.

As can be seen from Figure 6, the KD tree optimization algorithm proposed in this paper is also suitable for high-dimensional data, and the accuracy is also high, but the accuracy of the traditional algorithm is decreasing. Of course, when the data dimension is high, the time cost of the algorithm will increase accordingly.

The performance of the algorithm is analyzed from two aspects: data validity and algorithm running time. The evaluation of data validity focuses on the information loss caused by the algorithm after anonymity to the original social network. As shown in Figure 7, under the same degree

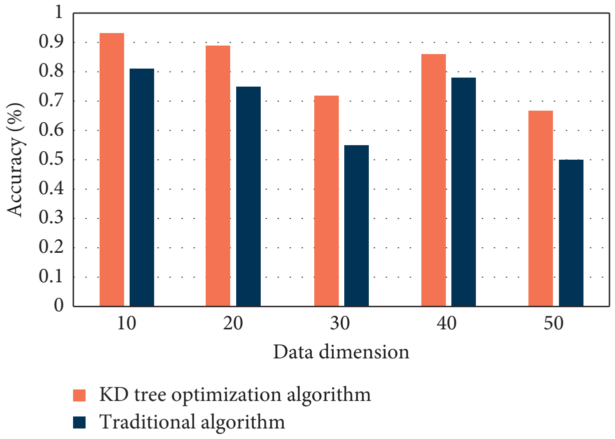

FIGURE 6: Accuracy analysis of experimental results in different dimensions.

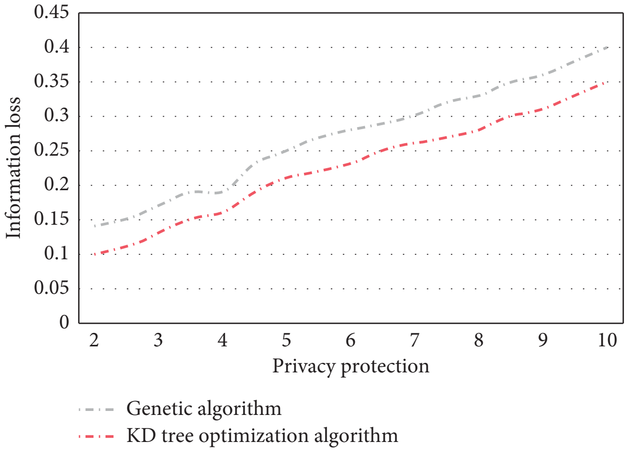

FIGURE 7: Comparison chart of algorithm information loss.

of privacy protection, the algorithm proposed in this paper has higher data availability.

As shown in Figure 8, under the same privacy protection strength, the KD tree algorithm takes less time, is more efficient, and has a higher time efficiency.

The algorithm in this paper first uses KD tree optimization to select k clustering centers. When there are new data, the nearest neighbor search method is used to cluster the new data reasonably so as to cluster the dynamic data quickly and efficiently. The algorithm only needs to process the incremental data so as to avoid reclustering all the data when the incremental data appears, thus improving the efficiency of clustering the incremental data to a certain extent. The increase in weight varies with k as shown in Figure 9, which basically changes linearly. The number of node splits varies with k as shown in Figure 10, which is positively correlated with the overall, but also related to the size of the data set. Because it needs to be affected by node grouping, if there are too many remaining nodes less than the value, too many nodes need to be split.

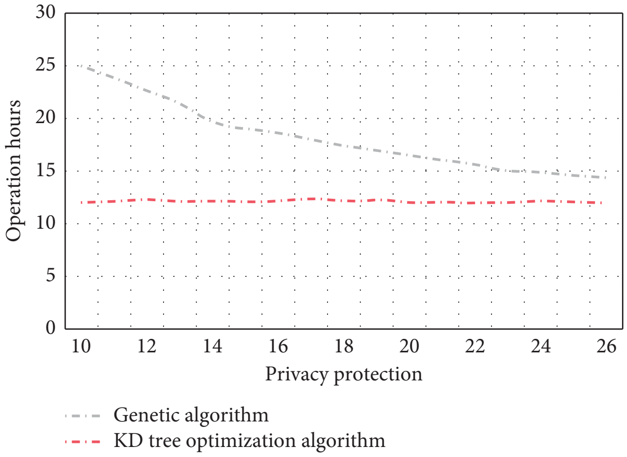

FIGURE 8: Comparison chart of the algorithm running time.

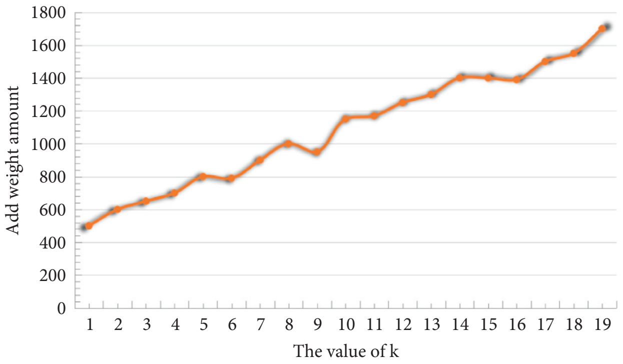

FIGURE 9: Schematic diagram of the change of the weight addition amount.

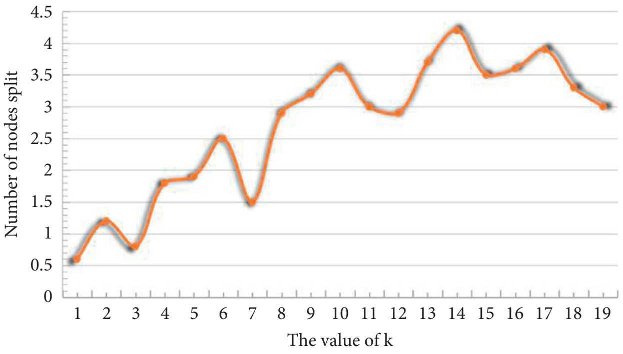

FIGURE 10: Schematic diagram of changes in the amount of node splitting.

There are many ways to protect users' privacy in social networks, but what we cannot ignore is how to ensure the practicability and availability of anonymized data. Anonymous social network graph should ensure that the user's identity is not identified and the user's sensitive information is not leaked. Although different applications may have different anonymous methods to process data, they should ensure the authenticity of the processed data, which has its due research and mining value when it is released. On the one hand, data mining of privacy data protection should take certain protection measures for privacy data; on the other hand, in data mining, the protected data is mined, and the algorithm adopted needs to eliminate and reconstruct the data, but the protection of privacy data that is damaged by decomposition will increase the storage capacity in decomposition, which wastes the storage space to a certain extent; at the same time, the damaged decomposition also effectively prevents data leakage and plays a very good security role. In the current era of big data, it is of great significance to explore the privacy protection of social networks.

5. Conclusions

With the continuous development of social network software and platform, a large number of data with social value and research significance have been accumulated. Data mining and analysis may lead to the disclosure of users' privacy. Therefore, how to ensure users' privacy security when effectively mining social networks is particularly important. Future research will mainly focus on the optimization of the algorithm so as to make it better applied to massive data. This paper proposes a KD tree optimal selection center point algorithm. Because it is easy to be attacked by external data in the process of dynamic clustering, the algorithm also introduces noise to disturb the data so as to achieve the effect of privacy protection. The data mining algorithm proposed in this paper can well protect the user's identity from being identified and the user's private information from being leaked. The algorithm can decompose the data and reconstruct the features and store the data vertically, which can effectively prevent the data from being threatened by security and will not cause the loss of mining accuracy. Only the processed data can be released to the public. Of course, while protecting the user's sensitive information, making the processed information still have certain usability is also an important factor in measuring data anonymity. The proposed anonymization algorithm is used in social networks to retain the structure of the original network and the effectiveness of the original data. While solving user identity authentication, data privacy disclosure, and information loss, it also protects the better application of social networks and creates more value. It is an important carrier for the development of the information age. Huge data is like a mine. The game between data mining and privacy protection is also a contest of technological development. Research on privacy protection technology of social network data still faces many new challenges, and there are still many problems to be further studied.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This paper was supported by the Research on Criminal Law Guarantee System of Network Security in China from the perspective of new development concept and overall national security concept (no. 21AZD082) and Research on Criminal Compliance Issues of Cross-Border e-Commerce Enterprises in Shaanxi Province under the background of "One Belt and One Road" (no.2021E003).

References

[1] P. Su, D. Yuan, and Martin, "Research on big data mining technology based on privacy protection," *Modern Computer: Professional Edition*, vol. 20, pp. 26–29, 2017.

[2] Z. Ali, M. Imran, S. Mcclean, N. Khan, and M. Shoaib, "Protection of records and data authentication based on secret shares and watermarking," *Future Generation Computer Systems*, vol. 98, pp. 331–341, 2019.

[3] F. Zhang and B. Shang, "The ethical dilemma and countermeasures of privacy protection in the era of big data," *People's Tribune Frontiers*, vol. 20, pp. 76–87, 2021.

[4] Y. Zhou, H. Chai, and Y. Zhao, "Analysis of the status quo and trends of big data research in the field of international library information," *Library Journal*, vol. 38, no. 12, pp. 18–29+46, 2019.

[5] W. Youke, W. Haiyang, W. Ningyun, and W. Yue, "An incentive-based protection and recovery strategy for secure big data in social networks," *Information Sciences*, vol. 508, pp. 79–91, 2020.

[6] N. A. Khan, S. Zhang, W. Zhou, A. Almogren, I. Ud Din, and M. Asif, "Inferring ties in social IoT using location-based networks and identification of hidden suspicious ties," *Scientific Programming*, vol. 2020, no. 1, pp. 1–16, 2020.

[7] B. Hu, "Implementation of data mining algorithms in big data security defense," *Industry and Technology Forum*, vol. 019, no. 7, pp. 48-49, 2020.

[8] Atiquzzaman, N. Yen, and Z. Xu, "Big data analytics for cyber-physical system in smart city," in *Proceedings of the BDCPS: International conference on Big Data Analytics for Cyber-Physical-Systems*, BDCPS, Shengyang, China, December 2019.

[9] W. Yan, "Research on the protection of women's rights and the countermeasures of social support based on big data network background," *International Journal for Engineering Modelling*, vol. 31, no. 1, pp. 252–257, 2018.

[10] M. Zhou, Z. Duan, and C. Shang, "Research on the three dimensions of education privacy protection in the big data era," *Guangxi Radio and TV University*, vol. 3, pp. 25–28, 2021.

[11] L. Zhang and B. Ashuri, "BIM log mining: discovering social networks," *Automation in Construction*, vol. 91, no. 7, pp. 31–43, 2018.

[12] G. Zuo, "Research on distributed privacy protection clustering mining algorithm based on big data," *Intelligent Computers and Applications*, vol. 8, no. 6, pp. 63–66, 2018.

[13] X. Zhang and Y. Wu, "Research progress of empirical asset pricing based on network big data mining," *Economic Trends*, vol. 000, no. 6, pp. 129–140, 2018.

[14] R. Yang, "Research on empirical asset pricing based on network big data mining," *National Circulation Economy*, vol. 2224, no. 28, pp. 69-70, 2019.

[15] X. Wang, "Artificial intelligence and privacy protection in the era of big data medical care," *Electronic Product World*, vol. 26, no. 6, pp. 84–86, 2019.

[16] R. Toujani and J. Akaichi, "Event news detection and citizens community structure for disaster management in social

networks," *Online Information Review*, vol. 43, no. 1, pp. 113–132, 2019.

[17] W. Yamin, Z. Fuanguo, W. Huaxiong, G. Zheng, M. Yinbin, and D. Yuqiao, "A new secret handshake scheme with multi-symptom intersection for mobile healthcare social networks-ScienceDirect," *Information Sciences*, vol. 520, pp. 142–154, 2020.

[18] P. Zhao, K. Bian, T. Zhao et al., "Understanding smartphone sensor and app data for enhancing the security of secret questions," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 552–565, 2017.

[19] J. Zhang, Y. Ma, and W. Xie, "Research on differential privacy protection for location big data," *Software Guide*, vol. 017, no. 11, pp. 206–208, 2018.

[20] X. Chen, "Research on social network data mining technology based on naive bayes algorithm," *Computer Measurement & Control*, vol. 25, no. 6, pp. 199–202, 2017.

[21] M. Hou, R. Wei, X. Lan, L. Xing, T. Na, and L. Lu, "Application research of medical big data privacy protection model based on differential privacy," *China Digital Medicine*, vol. 014, no. 12, pp. 86–88, 2019.

[22] K. Muhammad, M. Sajjad, I. Mehmood, S. Rho, and S. Baik, "Image steganography using uncorrelated color space and its application for security of visual contents in online social networks," *Future Generation Computer Systems*, vol. 86, no. 9, pp. 951–960, 2016.

[23] H. Chen, G. Wang, and P. Zhang, "Key nodes mining algorithm in Sina Weibo social network based on Hadoop cloud platform," *Dongnan Daxue Xuebao (Ziran Kexue Ban)/ Journal of Southeast University (Natural Science Edition)*, vol. 48, no. 4, pp. 590–595, 2018.

[24] P. Pinto, I. Theodoro, M. Arrais, and J. Oliveira, "Data mining and social web semantics: a case study on the use of hashtags and memes in Online Social Networks," *IEEE Latin America Transactions*, vol. 15, no. 12, pp. 2276–2281, 2017.

[25] A. Farasat, G. Gross, R. Nagi, and A. G. Nikolaev, "Social network analysis with data fusion," *IEEE Transactions on Computational Social Systems*, vol. 3, no. 2, pp. 1–12, 2016.

[26] G. Fanti, P. Kairouz, S. Oh, K. Ramchandran, and P. Viswanath, "Hiding the rumor source," *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6679–6713, 2017.

[27] H. Zhu, Y. Zhang, and Z. Yan, "A provably password authenticated key exchange scheme based on chaotic maps in different realm," *International Journal on Network Security*, vol. 18, no. 4, pp. 688–698, 2016.

[28] B. Martin, "\"Let's protest\": surprises in communicating against repression," *IEEE potentials*, vol. 35, no. 5, pp. 16–18, 2016.