Social Network Privacy Protection Based on Differential Privacy Technology and Community Discovery Algorithm

Xia Wu

Department of Information Engineering, Henan Vocational College of Water Conservancy and Environment, Zhengzhou 450008, China
E-mail: xiawu15@outlook.com

Received 05 January 2026; Accepted 13 April 2026

Abstract

The high aggregation of user relationship and behavioral data in social networks continues to aggravate privacy leaks. How to strike a balance between privacy protection and data availability has become a research hotspot. To collaboratively optimize user information security and community structure identification, this study proposes a social network privacy protection model that integrates differential privacy technology and community discovery algorithms. First, a differential privacy noise injection mechanism is constructed to perturb node data and combine it with blockchain storage to ensure that the data cannot be tampered with. Then, a community division strategy based on information entropy and mutual information is introduced to achieve high-precision community identification through modularity optimization. The accuracy of the proposed model reached 98.1% when the data set size was 800, which was about 3.4% and 9% higher than that of other models, respectively. The root mean square error was 8.2, which was about 20% lower than that of the traditional model. The convergence speed was increased to 380 iterations, which was about 15% faster than that of the comparison

Journal of Cyber Security and Mobility, Vol. 15_4, 799–822.
doi: 10.13052/jcsm2245-1439.1542

algorithm. The privacy protection strength and scalability scores reached 9.3 and 9.5, respectively. The simulation test results showed that, under different data types, the accuracy of the model grew from 0.87 to 0.98, and the F1 value grew from 0.84 to 0.95. The integration of differential privacy and community discovery effectively improves the privacy protection strength and structural analysis accuracy of social networks, providing a highly feasible solution for multi-scenario social data security analysis.

Keywords: Differential privacy, social network, community discovery algorithm, privacy protection model, data security.

1 Introduction

With the widespread popularity of Internet social platforms and the rapid development of intelligent terminal devices, human society is communicating and interacting in virtual space on an unprecedented scale. Social networks have not only become an important medium for individual information dissemination and behavioral expression, but also an important data source for social relationship modeling, group behavior prediction, and public opinion analysis. However, the high-correlation and high-dimensional structural characteristics of social data significantly increase privacy leakage. Users' identity information, behavior trajectories, interests and preferences, can easily be inferred and restored during the data sharing or mining process, thus causing privacy exposure and data abuse issues. Traditional privacy protection methods, like data anonymization, disguise processing and access control, can no longer take into account data availability and security when faced with complex graph structures and multi-source heterogeneous social data [1–3]. Excessive anonymization will lead to a significant decrease in data analysis accuracy and weaken the effectiveness of network structure research and community detection. In addition, insufficient protection will expose individual characteristics to differential reasoning or correlation analysis by external attackers. Although these studies are not specifically designed for social network analysis, they provide important theoretical foundations in privacy-preserving computation, secure communication protocols, and distributed data protection mechanisms. Their design principles contribute to the development of privacy-aware graph analytics frameworks. Xu et al. designed a safe and efficient privacy-protecting communication protocol to address the threats of information leakage and data attacks faced by smart city Internet of Vehicles. This protocol achieved higher security while

significantly reducing computing and communication overhead [4]. Ahmed et al. built a blockchain-based video sharing solution to ensure the safe and reliable dissemination of media content in video streaming applications. Smart contracts were used to manage data flows, and encryption primitives were used to protect the privacy of streaming media and viewers. Research showed that this solution effectively protected user privacy while ensuring the authenticity of multimedia [5]. Haipeng et al. developed a privacy-protecting cross-domain identity authentication protocol for the complex cross-domain operations and frequent terminal access and exit requirements in edge cloud collaboration scenarios, which had functions such as self-certified key generation, cross-domain authentication, and identity revocation. Experiments showed that this protocol was both efficient and practical while ensuring security [6]. To solve the privacy and access control issues in medical data sharing, Chen et al. integrated K-anonymity and searchable encryption technology to achieve secure data interaction between medical institutions and data users. Results showed that this scheme had good efficiency while meeting privacy protection and access control [7]. To enhance the accuracy of epilepsy brainwave signal detection, Zhang et al. built an unsupervised multi-view multi-class variable entropy fuzzy clustering method. Multi-angle feature extraction, multi-medoid cluster structure modeling, variable entropy collaborative learning and privacy protection mechanisms were applied to optimize the clustering effect. This method performed well in epilepsy signal detection and effectively protected user privacy [8]. However, the above studies still have poor protection performance and lengthy running times. This study proposes a social network privacy protection model based on Differential Privacy (DP) and community discovery algorithms. Unlike existing DP-based graph publishing or community detection methods, they usually inject uniform noise without structural awareness. The proposed framework introduces an entropy-guided privacy control mechanism and a mutual-information-driven structural reinforcement strategy after node perturbation. Specifically, information entropy is employed to identify structurally critical nodes before community division, and mutual information is used to preserve high-correlation node relationships under noise constraints. Furthermore, blockchain-based hash indexing is integrated to ensure data immutability and traceable storage, which is rarely considered in traditional DP-based community detection methods. Through the coordinated design of dynamic privacy budget allocation, structural feature preservation, and modularity optimization, the model achieves privacy protection while minimizing structural information loss. Therefore, the novelty lies in the

structurally aware integration of DP, information-theoretic feature screening, and secure storage mechanisms within a unified community discovery framework.

Compared with existing DP-based community detection methods, the proposed approach introduces three key distinctions. First, traditional DP-based graph publishing or community detection frameworks generally inject uniform noise into adjacency matrices or edge weights without structural awareness, which often leads to excessive distortion of local connectivity patterns. In contrast, the proposed model incorporates entropy-based node importance evaluation before community division, allowing privacy protection to be applied while preserving structurally informative nodes. Second, many existing methods rely solely on modularity optimization under noisy graphs, whereas this work integrates mutual-information-driven correlation reinforcement to mitigate topology disruption caused by noise injection. Third, previous DP-based approaches typically focus on either privacy-preserving graph release or community detection independently, while the present model combines dynamic privacy budget allocation, structural feature screening, and blockchain-assisted integrity verification within a unified framework. These differences enable a better balance between privacy guarantees and structural fidelity compared with conventional DP-based community detection techniques.

2 Methods

2.1 Social Network Privacy Protection Model Based on DP Technology

The high-density aggregation of user relationship and behavioral data in social networks has significantly increased the privacy leakage, making it difficult for traditional anonymization and encryption mechanisms to achieve an effective balance between data availability and security. DP, as a mathematical framework that protects individual information by adding random noise, provides a new solution for privacy protection in social networks [9]. The core idea is to introduce controlled perturbations in the process of statistical query or structural analysis so that external attackers cannot accurately infer the real information of a single node or edge, thereby achieving strict constraints on individual privacy while ensuring the effectiveness of global data features. The structure of the social network privacy protection model based on DP technology is presented in Figure 1.

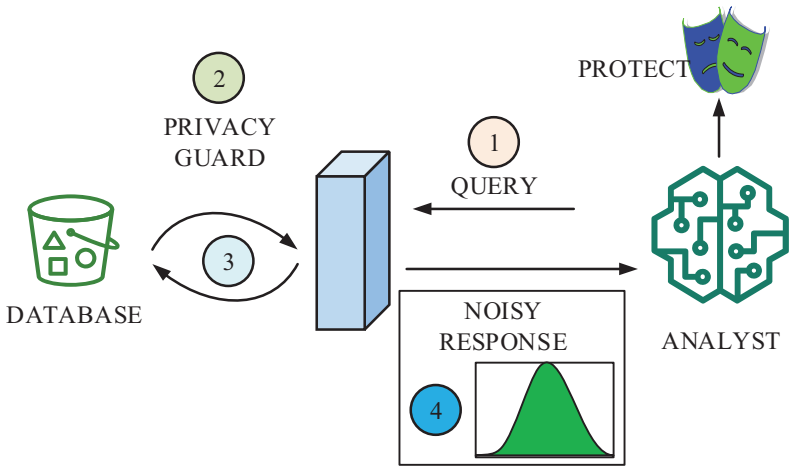

Figure 1 Social network privacy protection model structure based on DP technology.

From Figure 1, the analyst first sends a query request to the system, and its expression is shown in Equation (1).

$$Q(D) \rightarrow f(D) \tag{1}$$

where D denotes the original data set in the database and $f(D)$ denotes the query function output. To protect individual privacy, the system introduces noise into the output results through the privacy protection module [10]. Its core mechanism is to ensure that any two data sets D_1 and D_2 that differ by only one sample will have an approximate output distribution under the same query, and its expression is shown in Equation (2).

$$\Pr[M(D_1) \in S] \leq e^{\varepsilon} \Pr[M(D_2) \in S] \tag{2}$$

where $M(\bullet)$ represents the randomization algorithm, S is any output set and ε signifies the privacy budget. The smaller the value, the stronger the privacy protection. The privacy protection module implements perturbation by adding Laplacian noise or Gaussian noise, where the Laplacian mechanism is shown in Equation (3).

$$M(D) = f(D) + Lap\left(\frac{\Delta f}{\varepsilon}\right) \tag{3}$$

where Δf is the global sensitivity, which reflects the maximum impact of a single sample on the output result. The added noise obeys the Laplace

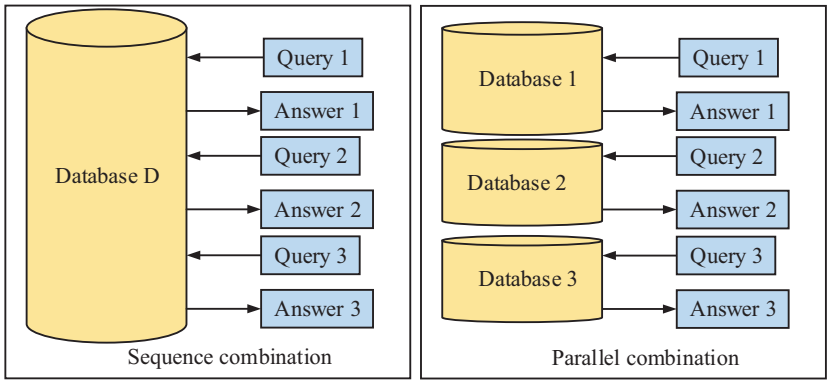

Figure 2 Parallel combination and sequence combination of DP.

distribution with mean 0 and scale $\frac{\Delta f}{\varepsilon}$. The Gaussian mechanism is shown in Equation (4).

$$M(D) = f(D) + N(0, \sigma^2) \tag{4}$$

where σ is the standard deviation, which controls the noise intensity. The entire process includes four steps, namely, the analyst issues a query, the privacy protection module receives the request and performs noise injection, the database provides real results to the privacy module, and outputs noisy response results [11, 12]. To optimize the performance, the DP mechanism further achieves privacy management in complex data scenarios through parallel combination and sequence combination. Its structure is shown in Figure 2.

From Figure 2, the sequence combination on the left indicates that multiple queries are applied to the same database in sequence. For example, the analyst continuously asks different questions, and the system returns corresponding noisy results, respectively. Since each query consumes a certain privacy budget, the overall privacy loss will gradually accumulate as the number of queries increases. Therefore, the query frequency is controlled reasonably or a smaller privacy budget is allocated for each operation in the sequence combination to prevent excessive leakage of data features. The parallel combination on the right indicates that different queries act on non-overlapping data subsets, and each subset independently implements the DP mechanism. In this way, the privacy risk of the overall system is no longer linearly superimposed but is determined by the largest privacy budget in each subset [13, 14]. Parallel combination is often used in multi-tasking or multi-data source scenarios, which can significantly improve computing

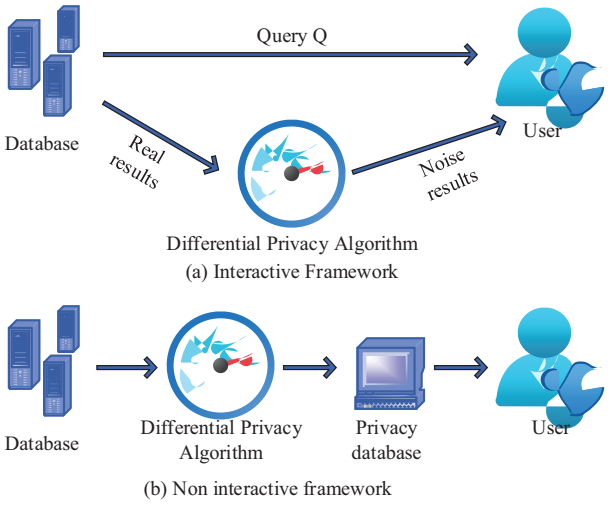

Figure 3 Work framework analysis.

efficiency and privacy control flexibility. Sequence combination is suitable for processing multiple analyzes of the same data set, emphasizing the cumulative characteristics of privacy budget. Parallel combination is suitable for partitioned or distributed data, reflecting independence and scalability. The combination of the mechanisms offers a theoretical basis for dynamic scheduling and policy optimization of DP in complex systems, allowing it to take into account data availability and privacy security in multi-query, multi-node and cross-domain computing environments. DP is mainly divided into two working frameworks: interactive and non-interactive. This study combines the two interactive forms to build the model. Its structure is shown in Figure 3.

From Figure 3, the upper half is an interactive frame and the lower half is a non-interactive frame. The interactive framework emphasizes the dynamic interaction between the user and the database. Every time the analyst makes a query request, the system immediately injects noise into the results through the privacy protection module and returns a perturbed answer. The analyst can then adjust the next query strategy based on the returned results. This process ensures the flexibility and real-time nature of queries but requires strict privacy budget management for each visit to prevent cumulative privacy leaks caused by multiple rounds of interactions [15, 16]. The non-interactive framework adopts a one-time data release strategy. The system first adds noise to the original data or performs aggregation processing before the data is

disclosed and generates statistical data or synthetic data sets protected by DP. Analysts can complete subsequent research and analysis without continuous interaction with the database. Although this method is not as flexible as the interactive framework, it can effectively control the overall privacy risk and is suitable for scenarios with high long-term security requirements such as data sharing and open statistical release. The two frameworks together constitute the core implementation model of DP under different application requirements, providing a systematic solution for balancing data availability and privacy protection.

2.2 Privacy Protection Model Based on DP Technology and Community Discovery Algorithm

In the first part, the research builds a privacy protection model based on DP technology. However, the performance of this model is not ideal when facing community data. Therefore, the study further combines the characteristics of social network data to build a comprehensive privacy protection model based on DP and community discovery algorithms. Social network data usually contain sensitive information such as user identity, interaction relationships, and behavioral trajectories. Its structure is complex and highly correlated. The current model is constructed based on static network snapshots. However, the framework can be extended to dynamic (temporal) social networks by introducing time-window segmentation or incremental update mechanisms. In such scenarios, DP noise can be applied to each temporal snapshot or to incremental graph changes, while entropy and mutual information calculations can be updated locally for evolving node neighborhoods. This design allows the model to preserve privacy guarantees while adapting to structural evolution over time. It is difficult for traditional anonymization or desensitization methods to effectively prevent correlation inference attacks while ensuring data availability. DP technology introduces controlled noise into data statistical results to ensure that the participation or absence of individual user information has minimal impact on the overall analysis results, thereby achieving strict privacy protection [17, 18]. The community discovery algorithm can identify potential community relationships in the graph structure, achieve information aggregation and local disturbance by dividing node clusters, and provide structural support for privacy protection. The algorithm steps are shown in Figure 4.

From Figure 4, first, the system performs DP encryption on the original social network data and reduces the risk of individual feature exposure

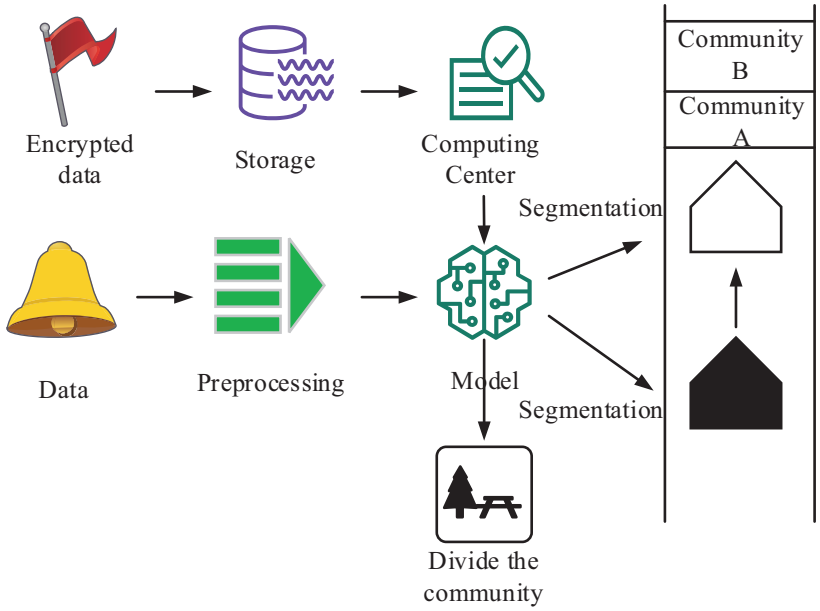

Figure 4 Community discovery algorithm steps.

through noise injection before the data is released. To control privacy loss more accurately, the privacy loss function is used to measure the cumulative risk of the algorithm under multiple rounds of operations, as shown in Equation (5).

$$\mathcal{L}(o) = \ln \frac{\Pr[M(D_1) = o]}{\Pr[M(D_2) = o]} \tag{5}$$

where $\mathcal{L}(o)$ denotes the privacy loss random variable, $M(D_1)$ and $M(D_2)$ represent the output results of executing the algorithm on adjacent data sets, respectively, and o refers to a possible output event. Secondly, the encrypted data is stored in a permissioned blockchain through a hash indexing mechanism to ensure non-tamperability of data transmission and the traceability of access. In this study, a consortium (permissioned) blockchain architecture is adopted instead of a public blockchain to reduce latency and enhance computational efficiency in multi-node collaborative environments. Only cryptographic hash values, access logs, and privacy budget records are stored on-chain, while perturbed social network data and community analysis results are maintained off-chain in a secure database. This hybrid on-chain/off-chain storage strategy minimizes storage overhead and prevents sensitive structural

data from public exposure, while still guaranteeing data integrity verification and auditability. Subsequently, the information entropy of the system node is used to determine the key nodes in the network, and its expression is shown in Equation (6) [19].

$$H(x) = -\sum_{i=1}^{n} p_i \log p_i \tag{6}$$

where p_i signifies the probability distribution of node characteristics or connections, which is used to evaluate the importance and information uncertainty of nodes. The correlation is calculated based on the mutual information between nodes to measure the similarity between nodes, as presented in Equation (7).

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{7}$$

where p_i signifies the joint probability distribution of nodes X and Y, and $p(x)$ and $p(y)$ signify marginal probabilities, respectively. Finally, the community is divided based on the mutual information matrix and node centrality results, and the modularity optimization function is shown in Equation (8).

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \tag{8}$$

where A_{ij} signifies the adjacency matrix element, k_i and k_j are node degrees, m signifies the total edges, and $\delta(c_i, c_j)$ is the community indicator function. This model achieves the security protection of social network data and efficient identification of community structures through privacy loss control, blockchain storage and modularity optimization. Its community social network diagram is shown in Figure 5.

From Figure 5, the upper part is the original social network graph, with nodes representing users or accounts, and edges representing interactive relationships between users, such as friend links, information exchange, or content forwarding. The nodes in the original network are densely connected and randomly distributed, with multi-level complex relationship structures. The intensity of interactions between different users varies greatly, which can easily lead to privacy leaks due to correlation reasoning during the analysis process. The lower part is the social network graph. The system divides the network into several relatively independent community structures by calculating the mutual information and centrality indicators between nodes. Each dotted circle represents a community. The nodes within the community

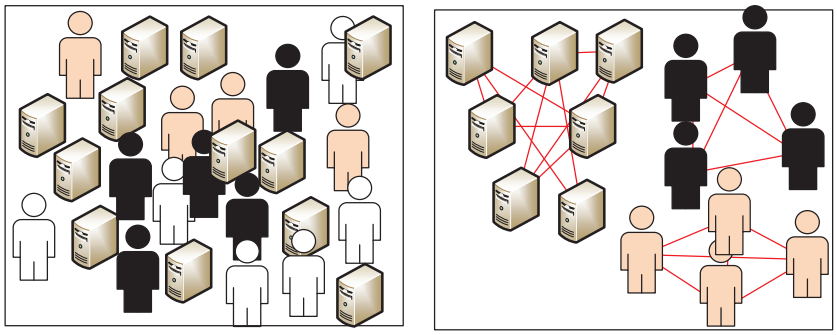

Figure 5 Social network diagram.

are closely connected and interact frequently, forming strong aggregation characteristics. The connections between different communities are relatively sparse, indicating weak cross-group interaction. Such a division can not only effectively reveal potential community relationships in social networks but also achieve partitioned privacy protection by introducing DP noise at the community level. It is important to clarify that community-level noise injection does not replace node-level DP but operates under the post-processing property of DP. Since DP guarantees are preserved under any deterministic or randomized post-processing that does not access the original raw data, the subsequent community aggregation and noise adjustment procedures do not increase privacy leakage risk. In this model, perturbation is first applied to node-level data before community detection is performed. Community-level operations are conducted only on already perturbed data, which means no additional information about individual nodes can be inferred beyond the privacy budget constraints. Therefore, the community-level noise redistribution does not violate node-level privacy guarantees, nor does it expose identifiable structural details of individual users. The model combines noise injection with community division in the analysis, so that the behavioral characteristics of individual nodes are blurred at the statistical level, thereby reducing the privacy leakage while maintaining the overall usability of the community structure. The final model process is shown in Figure 6.

From Figure 6, the algorithm is mainly divided into six stages. First, the Laplacian noise mechanism of DP is used to perturb node data to generate noisy data to ensure the privacy and security of individual node characteristics. The encrypted data is uploaded to the blockchain system to achieve distributed secure storage. The information entropy and probability density function of each node are calculated to obtain the importance index of the

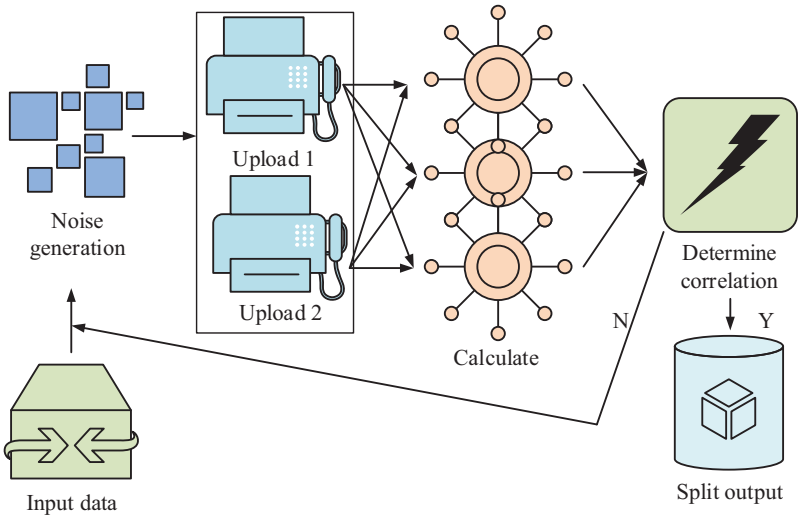

Figure 6 Social network privacy protection model process based on DP technology and community discovery algorithm.

node. Next, the central node set is determined according to the information entropy threshold. By calculating the mutual information between nodes, measuring the correlation between the node and the central node, the nodes with strong information correlation are divided into the same community. To ensure reproducibility and avoid subjective parameter setting, the entropy threshold used for central node selection is determined through an adaptive statistical strategy. After computing entropy values for all nodes, the distribution of entropy across the network is analyzed. Nodes whose entropy values exceed the network-wide average level and fall within the upper dispersion range are selected as candidate central nodes. Regarding mutual information computation in large-scale graphs, direct pairwise calculation among all nodes would result in excessive computational overhead. Therefore, mutual information is only computed between candidate central nodes and their local neighboring nodes within a limited hop range. In addition, sparse adjacency representations are used to reduce redundant probability estimation, and sampling-based approximation is applied for large neighborhoods. These strategies significantly reduce computational complexity while preserving structural correlation patterns. Finally, all central nodes and their associated nodes are traversed to obtain the final community set. The overall process integrates the security of DP and the structural identification capabilities of

information theory to achieve secure clustering and privacy protection in social networks.

3 Results

3.1 Performance of Privacy Protection Model Based on DP and Community Discovery Algorithm

This study used Windows 10 system, desktop computer with 16 GB RAM, equipped with Intel (R) Core (TM) i5-12600KF CPU, and GPU NVIDIA GeForce RTX 4090D. The data set adopts the NAP Facebook Social Circles public data set, which takes anonymized Facebook user social circles as the research object, and uses nodes to represent users and edges to represent friendship relations to build a social graph structure. The data comes from multiple "self-networks". Each self-network takes a certain user as the central node and records its direct friends and the interconnected relationships between friends. The data set not only contains social connection topology information, but also provides node attributes (such as educational background, work unit, city of residence) and social circle label information. These labels can be regarded as naturally formed community division results, providing a standard reference for verifying the accuracy and robustness of community discovery algorithms. The Facebook Social Circles data set has 4039 nodes and 88,234 edges, with an average clustering coefficient of about 0.605. It shows typical "small world" and "strong aggregation" characteristics and can truly reflect the complex structure of interpersonal relationships in social networks. Before conducting the experiments, several preprocessing steps are performed to ensure data consistency and experimental validity. First, isolated nodes and incomplete attribute records are removed to eliminate structural noise. Second, categorical node attributes are encoded using one-hot representation, while continuous attributes are normalized to avoid scale imbalance. Third, duplicate edges and self-loops are filtered to maintain a simple undirected graph structure. For experimental evaluation, the data set is partitioned using a stratified sampling strategy based on ground-truth social circle labels. A total of 70% of the nodes and their associated edges are used for community structure learning, while the remaining 30% are reserved for validation and performance testing. To evaluate scalability under different data sizes, subgraphs are randomly sampled from the original network while preserving local clustering characteristics. Each experiment is repeated five times with different random seeds, and the average results are reported to

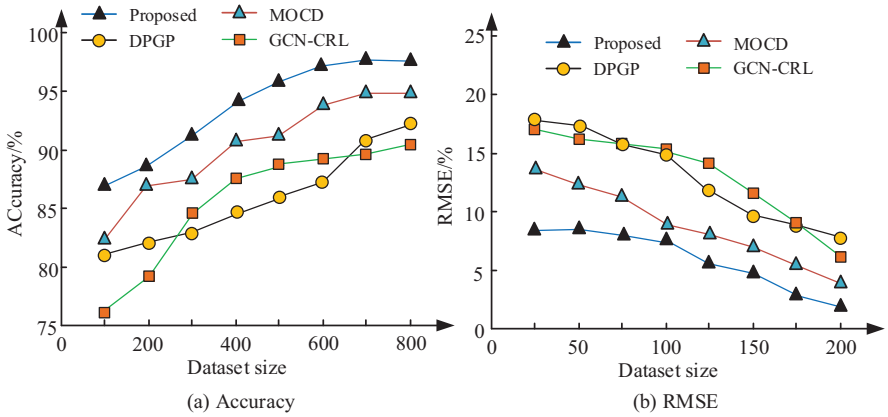

Figure 7 Comparison of accuracy and root mean square error of each model.

ensure statistical reliability. The study introduces the Differentially Private Graph Publishing Model (DPGP), the Modularity Optimization-Based Community Detection Algorithm (MOCD), and the Graph Convolutional Network for Community Representation Learning (GCN-CRL) as comparison models. The results are shown in Figure 7.

Figure 7(a) presents the accuracy change trend of different models when the data set size increases. Figure 7(b) presents the Root Mean Square Error (RMSE) error trend when the data set size changes. From Figure 7(a), as the data set size increased from 100 to 800, the accuracy of all models showed an upward trend. The proposed model performed the best, with an accuracy of about 98% when the data set size was 800, which was significantly higher than that of the MOCD (about 95%), DPGP model (about 91%) and GCN-CRL model (about 90%). This result shows that the proposed model has stronger feature learning and noise suppression capabilities under large-scale data. Its DP mechanism maintains high data availability while protecting user information, thus improving the community classification accuracy. The performance of MOCD is second, benefiting from the structural recognition ability of modularity optimization but, because no privacy protection is introduced, its generalization performance is slightly inferior to the proposed model. The robustness of DPGP and GCN-CRL under noise interference is relatively weak, especially the accuracy of DPGP under small sample conditions is less than 85%, indicating that the traditional perturbation mechanism causes a large loss of structural information. From Figure 7(b), the RMSE value of the proposed model was the lowest under all data sizes

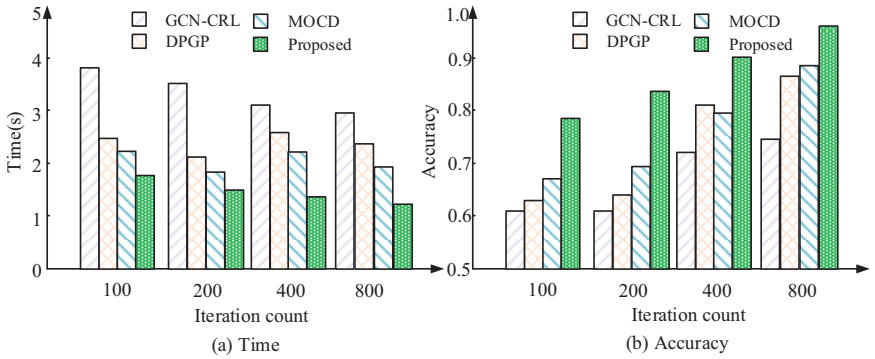

Figure 8 Trends in running time and accuracy of different models under different iterations.

and was only about 0.03 when the data set size reached 200. This shows that the proposed model maintains good stability and accuracy while introducing DP protection, and its error control ability is better than the comparison algorithm, verifying its efficiency and reliability in social network privacy protection scenarios. The performance under different iteration numbers was analyzed. The results are shown in Figure 8.

Figure 8(a) presents the running time change trend of different models under different iterations. Figure 8(b) presents the accuracy changes of different models under the same iterations. From Figure 8(a), as the iteration increased from 100 to 800, the running time of all models showed a downward trend. The proposed model took the shortest time, only about 1 second at 800 iterations, while MOCD was about 1.5 seconds, DPGP was about 2 seconds, and GCN-CRL still remained at about 2.5 seconds. From Figure 8(b), when the iteration reached 800, the accuracy of the proposed model was close to 1.0, which was significantly higher than that of the MOCD (0.92), GCN-CRL (0.90) and DPGP (0.83). This shows that the proposed model can reduce noise interference and improve the quality of node feature aggregation in multiple iterative learning, thereby more accurately identifying community structure. The accuracy of the DPGP model is always low because its noise injection strategy is too uniform, resulting in a large loss of local structural information. However, the proposed model achieves a balance between privacy protection and structure preservation through dynamic privacy budget allocation, showing better performance stability and robustness. The performance of each model was analyzed, as shown in Table 1.

From Table 1, the proposed model performed optimally in core indicators such as accuracy, error, convergence speed, privacy protection and robustness,

Table 1 Model comprehensive performance analysis

Index	DPGP	MOCD	GCN-CRL	Proposed Model
Accuracy/%	89.4	94.7	91.2	98.1
RMSE	14.8	10.3	11.7	8.2
Time/s	2.0	1.6	2.5	1.1
Modularity Q	0.72	0.81	0.78	0.88
Convergence rate/iter	600	450	520	380
Scalability/score	7.2	8.3	8.1	9.5
Privacy level/score	8.6	6.2	7.4	9.3
Robustness/%	86.3	90.1	88.7	95.9

showing significant comprehensive performance advantages. Specifically, the accuracy of the proposed model reached 98.1%, which was approximately 3.4% and nearly 9% higher than that of the MOCD (94.7%) and DPGP (89.4%), respectively, indicating that it had stronger feature expression and noise suppression capabilities in community segmentation accuracy. RMSE was 8.2, which was 2.1 lower than that of the MOCD, indicating that the prediction error was smaller and the model output result was more stable. Running time was only 1.1 seconds, which was nearly 45% shorter than that of the DPGP. Thanks to the parallel optimization design of information entropy screening and DP mechanism, computational complexity is effectively reduced. Convergence speed was 380 iterations, which was about 15% faster than that of the MOCD, reflecting higher training efficiency. The privacy protection strength and scalability scores were 9.3 and 9.5, which were better than those of other models, demonstrating that high-precision output can still be maintained under large-scale data and privacy constraints. The robustness reached 95.9%, which was nearly 10% higher than that of the DPGP. In this study, robustness is evaluated by combining three factors: resistance to noise perturbation, tolerance to partial data missing, and stability under simulated targeted inference attacks. Results indicate that the model maintains high structural consistency and low attack success probability even when adversarial nodes attempt to infer individual participation or attribute information from perturbed community outputs. The proposed model collaboratively optimizes accuracy, efficiency and security through dynamic privacy budget allocation and structural adaptation mechanisms and shows excellent overall performance in social network privacy protection and community identification tasks. In addition to overall performance comparison, the relationship between privacy strength and community detection quality was further analyzed. In the DP framework, the privacy budget ε directly

determines the trade-off between privacy protection and data utility. When ε decreases, stronger privacy protection is achieved due to higher noise intensity, but structural information distortion may increase. Experimental observations indicated that when ε was reduced from 1.0 to 0.2, privacy protection score increased by approximately 18%, while accuracy decreased from 98.1% to 94.6%, and RMSE increased from 8.2 to 10.7. Conversely, when ε was set above 1.2, accuracy improvement became marginal (less than 1.5%), while privacy strength significantly weakened. This demonstrates that excessive privacy constraints may deteriorate community boundary clarity, whereas overly relaxed privacy budgets reduce protection effectiveness. Therefore, a moderate ε interval (0.6–1.0) achieves a balanced trade-off between privacy strength and structural quality, ensuring stable modularity and acceptable error levels. The dynamic privacy budget allocation mechanism adopted in this study enables adaptive adjustment under different data scales, which mitigates the performance degradation commonly observed in fixed-budget DP-based graph models.

3.2 Simulation Performance Analysis

To further analyze the comprehensive performance, the study used simulation analysis to further verify the performance, as presented in Figure 9.

Figure 9(a) presents the accuracy under four data types. Figure 9(b) shows changes in the average RMSE values of different models under each data type. From Figure 9(a), as the complexity of the data type increased from Data A to Data D, the accuracy of all models showed an upward trend. The proposed model maintained the highest accuracy under each data type, which was approximately 0.87, 0.91, 0.94 and 0.95, respectively. From Figure 9(b), the overall trend was downward, indicating that the prediction error gradually

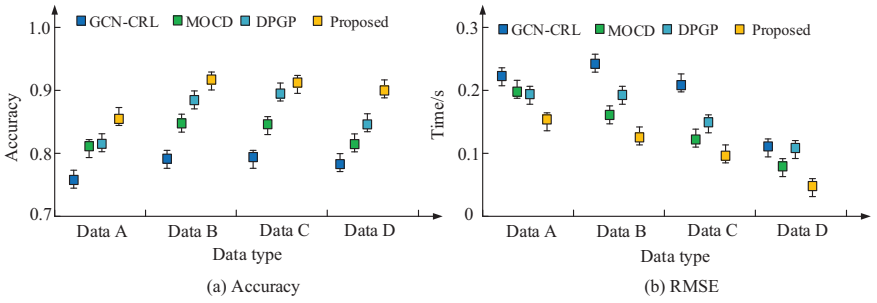

Figure 9 Performance trends of different models under four data types.

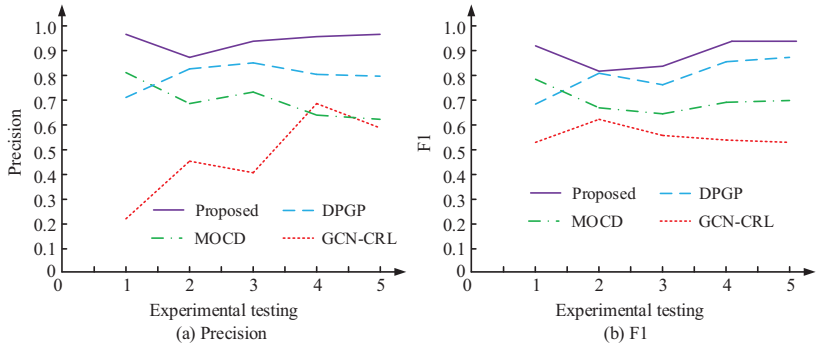

Figure 10 Change trends in precision of different models under five sets of experimental tests.

decreased after the data scale expanded. The proposed model had the lowest RMSE in the Data D stage, about 0.08, indicating that it had better error control and fitting capabilities. Five simulation tests were conducted, as presented in Figure 10.

Figure 10(a) shows the precision of different models under five sets of experimental tests. Figure 10(b) shows the changes in F1 values of different models. From Figure 10(b), the proposed model performed optimally in all tests, with the overall precision stable above 0.9, reaching a peak of approximately 0.98 in the 3rd and 5th tests, which was significantly higher than approximately 0.85 of DPGP, approximately 0.78 of MOCD, and approximately 0.45 of GCN-CRL. This shows that the model can still maintain strong community identification precision and feature consistency under privacy protection. From Figure 10(b), the F1 value of the proposed model remained above 0.8 and reached a peak of approximately 0.93 in the fourth test, which was significantly better than approximately 0.86 of DPGP, approximately 0.77 of MOCD, and approximately 0.58 of GCN-CRL, indicating that the model achieved a good balance between precision and recall. The proposed model improves the structural correlation between nodes through mutual information constraints, which reduces the misclassification rate and enhances the adaptability in complex networks. Overall, the proposed model shows the best stability and robustness in accuracy and comprehensive performance indicators. The simulation comprehensive performance of each model was analyzed, as presented in Table 2.

From Table 2, the proposed model performed best on all data sets. As the data complexity increased from Data A to Data D, the accuracy of the proposed model gradually increased from 0.87 to 0.98, F1 value increased

Table 2 Simulation performance analysis

Data	Model	Accuracy	F1	RMSE	Time/s	Modularity Q
Data A	Proposed	0.87	0.84	0.18	0.20	0.80
	MOCD	0.82	0.78	0.22	0.24	0.76
	DPGP	0.76	0.72	0.24	0.27	0.70
	GCN-CRL	0.72	0.67	0.26	0.30	0.68
Data B	Proposed	0.91	0.88	0.14	0.17	0.84
	MOCD	0.86	0.82	0.17	0.20	0.80
	DPGP	0.81	0.78	0.19	0.22	0.75
	GCN-CRL	0.79	0.75	0.20	0.24	0.73
Data C	Proposed	0.94	0.91	0.11	0.15	0.87
	MOCD	0.90	0.86	0.13	0.18	0.83
	DPGP	0.86	0.83	0.15	0.20	0.79
	GCN-CRL	0.84	0.80	0.16	0.22	0.78
Data D	Proposed	0.98	0.95	0.08	0.12	0.90
	MOCD	0.92	0.89	0.11	0.15	0.86
	DPGP	0.88	0.85	0.13	0.17	0.82
	GCN-CRL	0.86	0.82	0.14	0.19	0.81

from 0.84 to 0.95, and RMSE decreased from 0.18 to 0.08, reflecting the strong robustness and stability of the model under different data conditions. In comparison, the MOCD model had an accuracy of 0.92 and an F1 value of 0.89 in Data D. Although it has a high structural recognition ability, its adaptability under noise disturbance conditions is not as good as the proposed model. The accuracy of the DPGP in Data A to Data D increased from 0.76 to 0.88, with a small increase, indicating that its fixed noise mechanism still had information loss in complex social structures. GCN-CRL performed the weakest in multi-data scenarios. The accuracy of the Data D stage was only 0.86 and the RMSE was still as high as 0.14. The main reason is that its depth map convolution structure is sensitive to DP noise and has limited feature extraction capabilities. The proposed model is better than other algorithms in accuracy, error control, time efficiency and structural quality.

4 Conclusion

Aiming at the balancing privacy leakage risks and structural analysis accuracy in social networks, this study proposed a social network privacy protection model based on DP technology and community discovery algorithms. When the data set size was 800, model accuracy reached 98.1%, which was 3.4%

and 8.7% higher than that of MOCD (94.7%) and DPGP (89.4%), respectively. RMSE dropped to 8.2, and the error was about 30% lower than that of the GCN-CRL model (11.7). Running time was 1.1 seconds, which was 45% shorter than that of the DPGP. The modularity reached 0.88, which was about 8.6% higher than that of the MOCD (0.81), showing that the community division structure is tighter and the boundaries are clearer. The privacy protection strength and scalability scores were 9.3 and 9.5, respectively, which was about 15% higher than other models on average. Further simulation results showed that under four types of different complexity data, model accuracy increased from 0.87 to 0.98, F1 value increased from 0.84 to 0.95, RMSE decreased from 0.18 to 0.08, time consumption decreased from 0.20 seconds to 0.12 seconds, and modularity increased from 0.80 to 0.90. In contrast, the accuracy of the DPGP model in the Data D stage was only 0.88, MOCD was 0.92, and GCN-CRL was 0.86, which were all significantly lower than those of the proposed model. This model collaboratively enhances privacy security and analysis accuracy through dynamic privacy budget allocation and community structure optimization, can effectively resist correlation inference and differential attacks, and provides a feasible solution for social network data sharing and privacy computing. However, the current implementation is primarily designed for static network snapshots and may require further optimization for fully dynamic or high-frequency temporal social networks. Although the framework can theoretically be extended through time-window segmentation and incremental privacy budget management, further research is needed to design efficient temporal entropy updating and streaming mutual information estimation mechanisms. Future work will focus on adaptive privacy allocation across time slices and the integration of temporal graph learning strategies to enhance applicability in evolving social environments.

References

[1] Wu W, Zhang Y. An efficient intrusion detection method using federated transfer learning and support vector machine with privacy-preserving. Intelligent Data Analysis, 2023, 27(4):1121–1141. DOI:10.3233/IDA-226617.

[2] Xiao Z, Gao B, Huang X, Chen Z. An interpretable horizontal federated deep learning approach to improve short-term solar irradiance forecasting. Journal of Cleaner Production, 2024, 436(10):140585.1–140585.16. DOI:10.1016/j.jclepro.2024.140585.

[3] Fan M, Guo H. Privacy attack identification and protection strategy analysis based on vertical federation clustering. Journal of Cyber Security and Mobility, 2025, 14(2):475–504. DOI:10.13052/jcsm2245-1439.1429.

[4] Xu J, Li M, He Z, Tomley A. Security and privacy protection communication protocol for Internet of vehicles in smart cities. Computers and Electrical Engineering, 2023, 109(5):108778-1–108778-13. DOI:10.1016/j.compeleceng.2023.108778.

[5] Ahmed F, Wei L, Niu Y, Zhao T, Zhang W, Zhang D, Dong W. Toward fine-grained access control and privacy protection for video sharing in media convergence environment. International Journal of Intelligent Systems, 2022, 37(5):3025–3049. DOI:10.1002/int.22810.

[6] Haipeng S, Yu'An T, Congwu LI, Lei L, Zhang Q, Hu J. An edge-cloud collaborative cross-domain identity-based authentication protocol with privacy protection. Chinese Journal of Electronics, 2022, 31(4):721–731. DOI:10.1049/cje.2021.00.269.

[7] Chen Y, Meng L, Zhou H, Xue G. A blockchain-based medical data sharing mechanism with attribute-based access control and privacy protection. Wireless Communications and Mobile Computing, 2021, 21(5):1–12. DOI:10.1155/2021/6685762.

[8] Zhang Y, Jiang Y, Qi L, Bhuiyan M, Qian P. Epilepsy diagnosis using multi-view & multi-medoid entropy-based clustering with privacy protection. ACM Transactions on Internet Technology, 2021, 21(2):32–42. DOI:10.1145/3404893.

[9] Gao C, Yu J. Secure RC: A system for privacy-preserving relation classification using secure multi-party computation. Computers & Security, 2023, 128(12):341–357. DOI:10.1016/j.cose.2023.103142.

[10] Akbari-Nodehi H, Maddah-Ali MA. Secure coded multi-party computation for massive matrix operations. IEEE Transactions on Information Theory, 2021, 67(4):2379–2398. DOI:10.1109/TIT.2021.3050853.

[11] Lin Z, Zeng B, Huang Y, Yao Z, Hu H, Xu L. SASE: Self-adaptive noise distribution network for speech enhancement with federated learning using heterogeneous data. Knowledge-Based Systems, 2023, 266(22):1.1–1.15. DOI:10.1016/j.knosys.2023.110396.

[12] Cho YJ, Wang J, Chirvolu T, Joshi G. Communication-efficient and model-heterogeneous personalized federated learning via clustered knowledge transfer. IEEE Journal of Selected Topics in Signal Processing, 2023, 17(1):234–247. DOI:10.1109/JSTSP.2022.3231527.

[13] Han M, Xu K, Ma S, Li A, Jiang H. Federated learning-based trajectory prediction model with privacy preserving for intelligent vehicle. International Journal of Intelligent Systems, 2022, 37(12):10861–10869. DOI:10.1002/int.22987.

[14] Deebak BD, Hwang SO. Privacy preserving based on seamless authentication with provable key verification using mIoMT for B5G-enabled healthcare systems. IEEE Transactions on Services Computing, 2024, 17(3):1097–1113. DOI:10.1109/TSC.2024.3382950.

[15] Budati AK, Snv G, Cherukupalli K, Kumar A, Moorthy V. High speed data encryption technique with optimized memory-based RSA algorithm for communications. Circuit World, 2021, 47(3):269–273. DOI:10.1108/CW-10-2020-0282.

[16] Bhandari AK, Srinivas K, Kumar A. Optimized histogram computation model using cuckoo search for color image contrast distortion. Digital Signal Processing, 2021, 118(8):103–126. DOI:10.1016/j.dsp.2021.103203.

[17] Ratnayake H, Chen L, Ding X. A review of federated learning: taxonomy, privacy and future directions. Journal of Intelligent Information Systems, 2023, 61(3):117–125. DOI:10.1007/s10844-023-00797-x.

[18] Fang Y, Luo B, Zhao T, et al. ST-SIGMA: Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting. CAAI Transactions on Intelligence Technology, 2022, 7(4):744–757. DOI:10.1049/cit2.12145.

[19] Simon K, Vicent M, Addah K, Bamutura D, Atwiine B, Nanjebe D, Mukama AO. Comparison of deep learning techniques in detection of sickle cell disease. AIA, 2023, 1(4):252–259. DOI:10.47852/bonview AIA3202853.

Biography

Xia Wu (1979–), female, Han ethnicity, from Puyang, Henan, China, graduated with a bachelor's degree in Computer Science and Technology from North China University of Water Resources and Electric Power in 2004 and a master's degree in Control Engineering from Huazhong University of Science and Technology in 2008. From 2004 to present, she has been a full-time teacher at the School of Information Engineering, Henan Vocational College of Water Conservancy and Environment. Her research direction is computer application technology.