Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 18 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i18.51495

PAPER

# Privacy Protection in Mobile Big Data: Challenges and Solutions

**Peihua Su(✉)**

Institute of Technology,
Xi'an International University,
Xi'an, China

sphvsyj@163.com

**ABSTRACT**

The pervasive use of mobile big data has profoundly altered daily life, providing unprecedented convenience and efficiency. However, with the proliferation of mobile devices and the explosive growth of data volume, the issue of user privacy protection has become increasingly severe. Location information and semantic information, as the two core components of mobile big data, can directly reflect users' behavioral trajectories and thought dynamics, underscoring the importance of privacy protection. Although existing technologies can protect user data to a certain extent, traditional methods struggle to address increasingly sophisticated attack techniques in the face of evolving privacy threats. A comprehensive privacy protection scheme for mobile big data was proposed in this study, with a focus on two main areas: the privacy protection of location-based and semantic-based mobile big data. For location information protection, an uncertain graph model was employed to effectively resist combined attacks by jointly protecting the user layer and the location layer. For semantic information protection, a hypergraph clustering method was used to structurally protect the user layer and the semantic layer, enhancing the privacy security of semantic information. This study not only addresses existing gaps but also provides new solutions for mobile big data privacy protection, offering significant theoretical and practical value.

**KEYWORDS**

mobile big data, privacy protection, location information, semantic information, uncertain graph model, hypergraph clustering

## 1 INTRODUCTION

The extensive application of mobile big data has significantly transformed lifestyles, providing unprecedented convenience and efficiency. However, with the proliferation of mobile devices and the explosive growth of data volume, user privacy protection has become increasingly critical [1–4]. Location information and semantic information, as the two core components of mobile big data, can directly reflect users' behavioral trajectories and thought dynamics, underscoring the importance

Su, P.H. (2024). Privacy Protection in Mobile Big Data: Challenges and Solutions. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(18), pp. 49–61. https://doi.org/10.3991/ijim.v18i18.51495

Article submitted 2024-05-30. Revision uploaded 2024-07-09. Final acceptance 2024-07-29.

of privacy protection [5–7]. Although existing technologies can protect user data to a certain extent, traditional methods struggle to address increasingly sophisticated attack techniques amidst evolving privacy threats.

Protecting mobile big data privacy is not only a fundamental requirement for safeguarding personal information but also crucial for maintaining social trust and the healthy development of the data ecosystem [8, 9]. The leakage of location information may trigger severe security issues, such as tracking, surveillance, and location fraud. Conversely, the exposure of semantic information could lead to the revelation of thoughts, emotions, and sensitive data, potentially causing negative impacts on users' psychological well-being and social relationships [10–14]. Therefore, in-depth study and the development of innovative methods for mobile big data privacy protection are of paramount importance for ensuring data security and enhancing user trust.

Despite substantial study in the field of mobile big data privacy protection, the focus has primarily been on single-layer protection methods, such as protecting either location data or semantic data alone [15, 16]. These approaches often overlook the threat posed by attackers who may employ combined attacks using multi-layer data, resulting in limited privacy protection effectiveness [17–20]. Furthermore, existing studies largely concentrate on specific scenarios, lacking comprehensive protection strategies for different types of mobile big data, thereby falling short in addressing complex privacy attacks holistically.

To address these issues, a comprehensive privacy protection scheme for mobile big data was proposed in this study, focusing on two main areas: the privacy protection of location-based and semantic-based mobile big data. For location information protection, an uncertain graph model was employed to effectively resist combined attacks by jointly protecting the user layer and the location layer. For semantic information protection, a hypergraph clustering method was utilized to structurally protect the user layer and the semantic layer, enhancing the privacy security of semantic information. This study not only addresses existing gaps but also provides new solutions for mobile big data privacy protection, offering significant theoretical and practical value.

## 2 PRIVACY PROTECTION OF LOCATION-BASED MOBILE BIG DATA

In the current digital era, mobile big data has become an indispensable part of daily life. Mobile big data includes various data generated by users through mobile devices, with location-based and semantic-based data being particularly important. Location-based mobile big data records users' geographical positions and movement trajectories. This type of data not only exposes users' real-time locations but also reveals their travel patterns and lifestyle habits. If such data is exploited by malicious entities, users' physical safety may face significant threats. Additionally, when combined with other data, such as social network data, it is possible to infer users' identities, social relationships, and personal preferences. Conversely, semantic-based mobile big data includes users' text records, comments, and other semantic content. With the proliferation of social media platforms, the volume of text information shared by users on these platforms has increased dramatically. This semantic data reflects users' thoughts, emotions, and opinions and may contain sensitive personal information. If attackers employ combined attacks by leveraging both social network layers and semantic layers, the risk of privacy breaches for users will be significantly heightened. Therefore, researching privacy protection for these two

types of data is essential not only to prevent potential security threats but also to safeguard users' privacy rights and trust. This section elaborates on the privacy protection methods for location-based mobile big data.

## 2.1 Location hypergraph clustering method for mobile networks

In mobile networks, users associate themselves with their actual locations through check-in behaviors, forming a user-location mobile network graph, typically represented as a bipartite graph. Figure 1 illustrates the bipartite graph structure of the user-location network. In this figure, one set of nodes represents users, while the other set represents locations, with edges indicating users' check-in behaviors at certain locations. This graph is inadequate for directly characterizing commonalities among users and the overall network structure, posing challenges for understanding and analyzing user behavior patterns and location associations in mobile networks. To address these issues, the concept of a hypergraph was introduced in this study, along with a method for converting the bipartite graph into a hypergraph. A hypergraph is a more complex graph structure where a hyperedge can connect multiple nodes, thereby naturally representing complex multi-node relationships. Specifically, let $F = \{m_1, m_2, ..., m_v\}$ be the set of locations, with each element representing a specific place. The set of edges $I = \{I_1, I_2, ..., I_i\}$ of the hypergraph contains hyperedges, where each hyperedge $I_u = \{i_{u1}, i_{u2}, ..., i_{uj}\}$ includes multiple nodes, representing the associations among these nodes. In the context of the location mobile network applied in this study, the set of locations $F$ remains unchanged. The hyperedge rule is as follows: a hyperedge $I_{uk}$ includes nodes that are the common locations where users $I_u$ and $I_k$ have checked in within the bipartite graph of the location mobile network. The network model structure is given by the following equation:

$$I_{uk} = \left\{ m_j \mid (m_j, i_u) \in Z \Lambda (m_j, i_u) \in Z, \exists m_j \in F, i_u, i_k \in L \right\} \text{ and } I = I_{u \neq k} I_{uk} \tag{1}$$

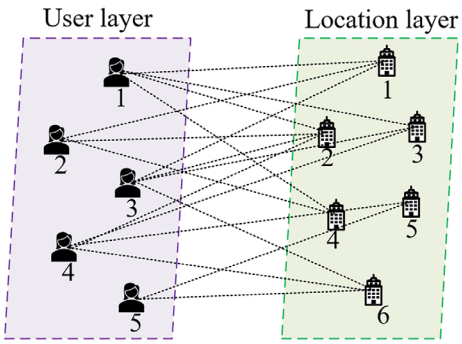


**Fig. 1.** Bipartite graph structure of the user-location network

In the hypergraph clustering method for mobile networks, the essence of the hypergraph structure lies in representing the associations among users and the distribution characteristics of locations in a more natural and intuitive manner. Traditional bipartite graphs classify users and locations as two distinct types of nodes, with edges indicating users' check-in behaviors at specific locations. Although this representation is straightforward, it fails to effectively capture and display the commonalities and complex associations among users. The hypergraph structure offers a new perspective by representing locations as vertices and using closed loops

to denote hyperedges that signify user check-in relationships. In this structure, each hyperedge encompasses multiple users' common check-in locations, thereby clearly illustrating shared check-in behaviors among users. This transformation enables the original bipartite graph, based on users and locations, to more intuitively reflect the associations among users and the overall check-in patterns. The association between any two hyperedges $I_a$ and $I_b$ in hypergraph $G$ can be calculated using the following equation:

$$SIM(I_a, I_b) = \frac{\left|I_a \cap I_b\right|^2}{\left|I_a\right|\left|I_b\right|} \tag{2}$$

Assuming the number of clusters is represented by $h$, and the partitioned clusters are denoted by the set $Z = \{Z_1, Z_2, ..., Z_h\}$. The objective function characterizing the similar hyperedge clustering problem was constructed to obtain the structural features of the network model, as shown in the following equation:

$$OBJ = \underset{Z}{\operatorname{argmax}} \sum_{h}^{s=1} \sum_{I_{uk} \in I \wedge I_{ik} \notin Z_S} SIM'(I_{uk}, Z_s) \tag{3}$$

The similarity $SIM'(I_{uk}, Z_s)$ between hyperedge $I_{uk}$ and cluster $Z_u$ can be calculated using the following equation:

$$SIM'(I_{uk}, Z_s) = \frac{1}{\left|Z_s\right|} \sum_{I_{ab} \in Z_s} SIM(I_{uk}, U_{ab}) \tag{4}$$

## 2.2 Privacy protection method for location-based mobile big data using hypergraph clustering

The proposed hypergraph clustering method for mobile networks transforms the relationships between users and locations into common behaviors among users and introduces an uncertain graph privacy protection mechanism within communities, providing an effective privacy protection strategy.

Specifically, hypergraph clustering assigns users to different clusters, with the set of clusters represented as $Z = \{Z_1, Z_2, ..., Z_h\}$. Each cluster $Z_u$ contains several users, i.e., $Z_1 = \{i_1, i_2, ..., i_v\}$. In the initial bipartite graph, each user is associated with several locations. Through hypergraph clustering, these associated locations of users can be clearly assigned to each cluster. For instance, the set of locations $\{m_1, m_2, m_3, m_4, m_6\}$ corresponds to $Z_1$, while the set of locations $\{m_2, m_3, m_5, m_6\}$ corresponds to $Z_2$. Figure 2 provides an example of the location hypergraph clustering process.

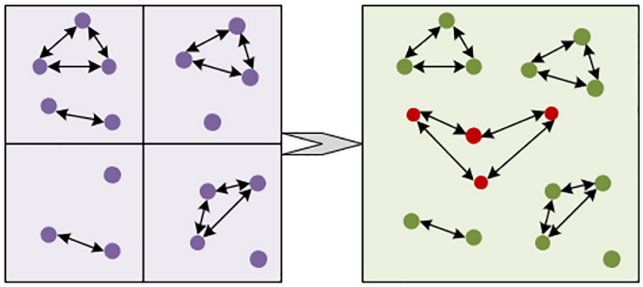

**Fig. 2.** Example of the location hypergraph clustering process

To achieve location privacy protection in mobile networks, an uncertain graph privacy protection algorithm was introduced in this study. Specifically, within each partitioned community, an edge was assigned between each user and each location within the community, with a probability assigned to each edge. This method introduces uncertainty, thereby obscuring the direct associations between users and locations to achieve privacy protection.

The use of uncertain graphs is crucial to the implementation of privacy protection. It allows for the generation of edges with probabilities within the community, where the probability represents the likelihood of a user checking in at a particular location. For example, in community $Z_1$, an edge between user $i_1$ and location $m_1$ is assigned a probability value. This design not only conceals the specific movement trajectories of users but also enhances the protection of location data, preventing malicious attackers from inferring specific user behaviors through simple graph structures.

## 3 PRIVACY PROTECTION OF SEMANTIC-BASED MOBILE BIG DATA

### 3.1 Semantic hypergraph clustering method for mobile networks

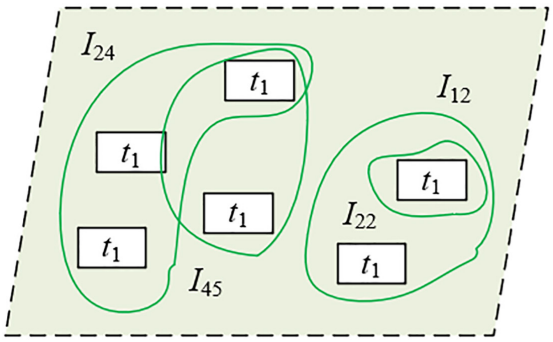


**Fig. 3.** Hypergraph representation of the user-semantics social network

In mobile networks, users associate with semantics through their commenting behaviors, forming a user-semantics mobile network graph. However, traditional bipartite graph structures only represent the single corresponding relationships between user nodes and semantic nodes, failing to capture the commonalities among users and the complex structural characteristics of the network. To represent clearly the intricate relationships among users in terms of semantics and to enhance the accuracy of data analysis and the effectiveness of privacy protection, the concept of hypergraphs was introduced in this study, converting the bipartite graph of the user-semantics mobile network into a hypergraph structure. Figure 3 illustrates the hypergraph representation of the user-semantics social network.

According to the definition of hypergraph $L$, let $T = \{t_1, t_2, ..., t_v\}$ be a finite set representing all the comment objects. The set $I = \{I_1, I_2, ..., I_i\}$ is the set of edges in the hypergraph, where each hyperedge $I_u$ contains multiple vertices, representing the associations among user nodes. Specifically, the set $T$ remains unchanged, representing the objects of user comments, while the hyperedge $I_{uk}$ represents the intersection of users $I_u$ and $I_k$ in the comment objects within the bipartite graph.

Under this transformation rule, the process of converting the bipartite graph of the user-semantics mobile network into a hypergraph is as follows: (a) The vertex set

remains unchanged. Firstly, all vertices in the user comment object set $T$ are retained. These vertices represent all the commented objects. b) Hyperedges are constructed. In the bipartite graph, if users $I_u$ and $I_k$ intersect on the same comment object, a hyperedge $I_{uk}$ is created in the hypergraph to connect these user nodes. The hyperedge includes multiple user nodes, indicating these users' common commenting behavior on certain objects. The transformation rule is characterized by the following equation:

$$I_{uk} = \left\{ t_j \mid (t_j, i_u) \in Z \Lambda (t_j, i_k) \in Z, \exists t_j \in T, i_u, i_k \in V \right\} \text{ and } I = I_{u \neq k} I_{uk} \tag{5}$$

In mobile networks, the semantic hypergraph clustering method achieves complex associations between users and semantics through the hypergraph structure, thereby improving the accuracy of data analysis and enhancing privacy protection. The hypergraph uses comment objects as vertices, and the hyperedges are composed of closed loops representing associated users. Each hyperedge includes the objects commonly commented on by users within that hyperedge, thus clearly presenting the relationship between users and comment objects in the graph structure.

Similar to the location hypergraph clustering method, the semantic hypergraph clustering method clusters user commenting behaviors by computing the hyperedge similarity and the hyperedge-cluster similarity, thereby enhancing the efficiency of data analysis and the effectiveness of privacy protection. Specifically, the similarity between any two hyperedges $I_a$ and $I_b$ can be calculated using a specific similarity measurement equation (Equation 2). To better compute similar structural features, the problem of clustering similar hyperedges was modeled as an objective function in this study, which achieves optimal clustering by maximizing the similarity of hyperedges within clusters. The similarity between hyperedge $I_{uk}$ and cluster $Z_u$ can be calculated using another equation (Equation 4), which clusters hyperedges based on their similarity, ensuring that hyperedges with high similarity are grouped into the same cluster, thereby maximizing the similarity value within the same community cluster.

## 3.2 Privacy protection method for semantic-based mobile big data using hypergraph clustering

In modern large-scale semantic mobile network graphs, traditional methods often focus on protecting the privacy of information contained within semantic nodes while neglecting the privacy of the network structure itself. This oversight can lead to sensitive information being revealed through the analysis of the network structure, even if the node information is protected. To address this issue, a privacy protection algorithm for semantic mobile networks based on hypergraph clustering was proposed in this study. This algorithm aims to protect the overall structure of semantic mobile networks, thereby enhancing data privacy while ensuring data usability. The algorithm operates based on the hypergraph clustering method, dividing mobile network users into several clusters, each considered a community, and applying an uncertain graph privacy protection mechanism within the communities. The specific steps are as follows:

Step 1: Parameter calculation. The algorithm first calculates the parameters $y$ required for the Laplace noise addition mechanism based on the global sensitivity $\Delta d$ and the privacy threshold $\gamma$. This step is crucial as it determines the intensity of the noise added subsequently, thereby affecting the effectiveness of privacy protection.

Step 2: Community division. Users and their comment objects are efficiently clustered to achieve privacy protection for semantic-based mobile big data.

This method first divides mobile network users into several clusters $Z = \{Z_1, Z_2, ..., Z_h\}$, each containing several users, facilitating the treatment of each cluster as a community. Users within these communities typically share similar comment objects, reflecting a certain semantic association. These clusters represent the natural grouping of users and comment objects, aiding in more effective local privacy protection.

Step 3: Noise addition. Within each community, the associations between users and comment objects can be handled using the uncertain graph privacy protection algorithm. This algorithm assigns an edge between users and comment objects within each community and assigns a probability value to this edge, thereby achieving uncertain graph data publishing for the local mobile network. Furthermore, the Laplace noise addition mechanism is used to add noise between any user and any comment object to obscure these associations, preventing external attackers from inferring user behaviors through these links.

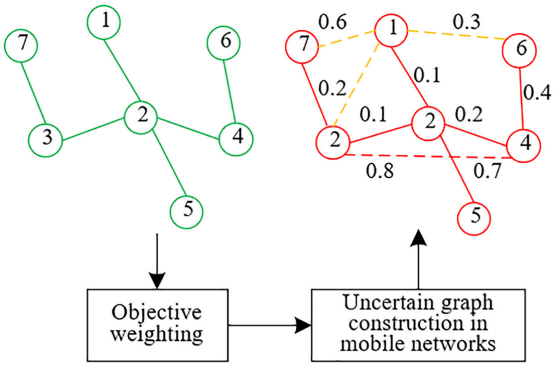


**Fig. 4.** Uncertain graph construction model of the semantic mobile network

Step 4: Noise transformation. The Laplace-generated noise is transformed into probabilities, which are then added to the edges between user nodes and comment object nodes. This step quantifies the impact of the noise into specific probability values, ensuring that the generated uncertain graph not only protects privacy but also retains partial data usability.

Step 5: Returning the uncertain graph. Finally, the algorithm returns the generated uncertain graph of the semantic mobile network. This graph, while maintaining the original network structure, effectively protects the privacy of the relationships between users and comment objects by adding noise. Figure 4 illustrates the architecture for the uncertain graph construction model of the semantic mobile network.

## 4 EXPERIMENTAL RESULTS AND ANALYSIS

As shown in Table 1, with the increase in the number of nodes, significant differences are observed between the original graph and the proposed method in terms of the number of edges, average degree of nodes, and node degree variance. For instance, when the number of nodes is 50, the original graph has 135 edges, an average degree of 5.02, and a degree variance of 60.25. In contrast, the proposed method shows a substantial increase, with 1254.21 edges, an average degree of 24.21, and a degree variance of 255.26. This trend becomes more pronounced with 100, 200, and 500 nodes. Under these conditions, the proposed method exhibits 7654.21, 22356.21, and 145216.25 edges, respectively, with significantly higher average degrees and degree variances than the original graph. Particularly when

the number of nodes reaches 500, the degree variance of the proposed method reaches 63215.25. These results indicate that the proposed method maintains a high number of edges and degree variance when handling larger-scale data. It is evident that the proposed method, through the uncertain graph model and the hypergraph clustering method, significantly increases the number of edges and the degree variance of the network graph. This suggests that while protecting privacy, the proposed method can substantially enhance the complexity and diversity of the data, thereby resisting potential privacy attacks. Higher average degrees and degree variances imply a more complex network structure, making it more challenging for attackers to deduce the privacy of other nodes from information on a single node.

**Table 1.** Data usability analysis of the original graph and the proposed method

| Metric \ Number of Nodes | Original Graph | | | Proposed Method | | |
|---|---|---|---|---|---|---|
| | Number of Edges | Average Degree of Nodes | Node Degree Variance | Number of Edges | Average Degree of Nodes | Node Degree Variance |
| 50 | 135 | 5.02 | 60.25 | 1254.21 | 24.21 | 255.26 |
| 100 | 389 | 9.62 | 198.63 | 7654.21 | 66.32 | 1548.23 |
| 200 | 1242 | 13.54 | 458.23 | 22356.21 | 154.25 | 5023.45 |
| 500 | 4521 | 20.35 | 1452.12 | 145216.25 | 326.25 | 63215.25 |

**Table 2.** Changes in metric values of the proposed method under different thresholds

| Metric \ Threshold | 0.01 | | |
|---|---|---|---|
| | Number of Edges | Average Degree of Nodes | Node Degree Variance |
| 50 | 1156.32 | 22.32 | 254.23 |
| 100 | 7548.26 | 60.21 | 1489.23 |
| 200 | 19623.25 | 151.24 | 4895.32 |
| 500 | 135695.26 | 302.32 | 61254.23 |
| **Metric \ Threshold** | **0.1** | | |
| | **Number of Edges** | **Average Degree of Nodes** | **Node Degree Variance** |
| 50 | 1089.23 | 19.87 | 245.23 |
| 100 | 7525.23 | 57.26 | 1468.23 |
| 200 | 19124.21 | 145.23 | 4569.23 |
| 500 | 13124.25 | 278.23 | 58954.26 |
| **Metric \ Threshold** | **1** | | |
| | **Number of Edges** | **Average Degree of Nodes** | **Node Degree Variance** |
| 50 | 986.21 | 18.26 | 236.24 |
| 100 | 7452.21 | 53.26 | 1356.87 |
| 200 | 18563.23 | 135.26 | 4325.21 |
| 500 | 122548.23 | 289.66 | 54879.26 |

The data in Table 2 indicates that as the threshold increases, there is a noticeable decline in the number of edges, average degree of nodes, and node degree variance using the proposed method. For example, when the threshold is 0.01, the number of nodes is 50, the number of edges is 1156.32, the average degree is 22.32, and the degree

variance is 254.23. As the threshold increases to 0.1 and 1, these metrics decrease to 1089.23 and 986.21 (number of edges), 19.87 and 18.26 (average degree), and 245.23 and 236.24 (degree variance), respectively. This trend is also evident with larger numbers of nodes (e.g., 100, 200, and 500), where an increase in the threshold leads to a reduction in the number of edges, average degree, and degree variance. For instance, when the number of nodes is 500, increasing the threshold from 0.01 to 1 result in the number of edges dropping from 135695.26 to 122548.23, the average degree from 302.32 to 289.66, and the degree variance from 61254.23 to 54879.26. The experimental results demonstrate that the performance of the proposed method under different thresholds shows that with an increase in the threshold, the complexity of the graph decreases. However, even at higher thresholds, the proposed method still maintains a high level of complexity and diversity. This indicates that the proposed method can flexibly balance privacy protection and data usability. Lower thresholds enhance the strength of privacy protection while increasing the complexity and diversity of the data; higher thresholds appropriately reduce the intensity of privacy protection but still maintain reasonable data complexity. This flexibility makes the proposed privacy protection scheme adaptable to different application needs and security requirements, offering high practicality and reliability in real-world applications.

**Table 3.** Changes in degree centrality edge entropy values of the proposed method under different thresholds

| Threshold Privacy Budget | 0.01 | 0.1 | 1 |
|---|---|---|---|
| 50 | 2561.21 | 2451.26 | 2451.23 |
| 100 | 10898.26 | 10785.23 | 10548.26 |
| 200 | 45265.27 | 45125.87 | 43926.25 |
| 500 | 122458.45 | 126895.25 | 122535.23 |

The data in Table 3 reveals significant differences in the changes in edge betweenness centrality values of the proposed method under various privacy budgets and thresholds. When the privacy budget is 50, the edge betweenness centrality values are 2561.21 (threshold of 0.01), 2451.26 (threshold of 0.1), and 2451.23 (threshold of 1). As the privacy budget increases to 100, these values rise to 10898.26, 10785.23, and 10548.26, respectively. With a further increase in the privacy budget to 200, the values continue to ascend to 45265.27, 45125.87, and 43926.25. At a privacy budget of 500, the edge betweenness centrality values reach their highest, at 122458.45, 126895.25, and 122535.23. These results indicate that as the privacy budget increases, the edge betweenness centrality values significantly rise, while the values also vary across different thresholds. The experimental results demonstrate that the proposed privacy protection scheme's edge betweenness centrality values increase significantly with the privacy budget, suggesting that as the privacy budget grows, the amount of data processed and protected by the model also increases, leading to higher edge betweenness centrality values. Additionally, the privacy budget threshold impacts the edge betweenness centrality values: lower privacy budget thresholds typically result in higher values, while higher thresholds lead to slightly lower values. Lower thresholds require stricter protection measures, resulting in more data processing and protection operations. Despite the significant increase in edge betweenness centrality values under high privacy budgets, the proposed method demonstrates its comprehensive advantages in protection effectiveness and processing capability, making it highly adaptable and practical in various application scenarios.

The data in Figure 5 demonstrates significant differences in the privacy protection effectiveness of different methods under varying user densities. The privacy protection evaluation metric for the local differential privacy method increases from 72% at a user density of 5 to 80% at a user density of 10. The metric for the differential privacy method based on the Laplace mechanism rises from 82.50% (user density of 5) to 87% (user density of 10). In contrast, the proposed method outperforms the other two methods at all user densities, from 92.50% at a user density of 5 to 96.50% at a user density of 10. These results indicate that as user density increases, the privacy protection effectiveness of all methods improves. However, the proposed method consistently maintains the highest privacy protection evaluation metric. The experimental results show that the proposed privacy protection scheme exhibits superior privacy protection capabilities across different user densities, significantly outperforming local differential privacy and the differential privacy method based on the Laplace mechanism. As user density increases, the privacy protection effectiveness of all methods improves, likely due to the higher aggregation of trajectory data enhancing the data perturbation effect. Notably, the proposed method, by combining the uncertain graph model and the hypergraph clustering method, effectively improves the privacy protection level of both location and semantic information, countering combined attacks and maintaining a high privacy protection evaluation metric under all user density conditions.

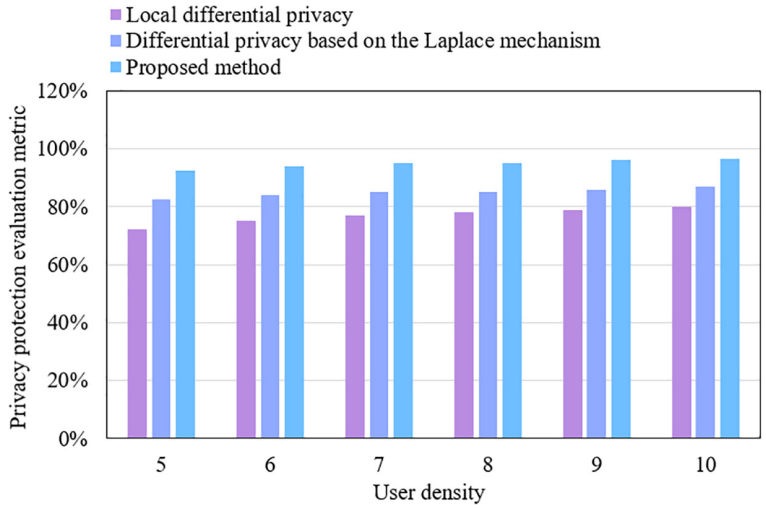


**Fig. 5.** Changes in the privacy protection evaluation metrics of different methods under varying user densities

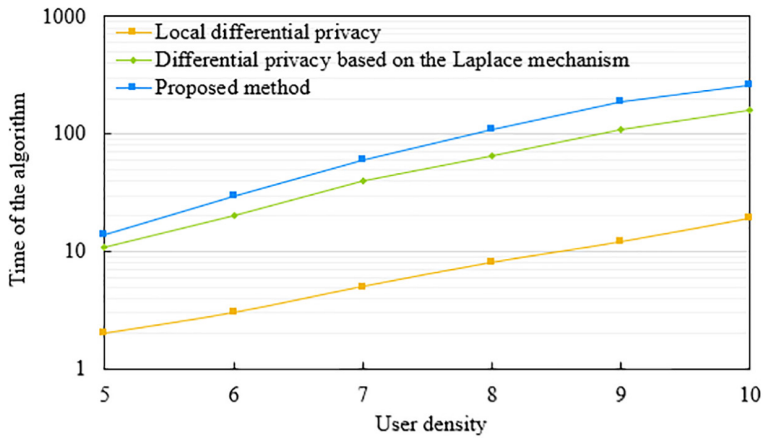


**Fig. 6.** Changes in the running time of different methods under varying user densities

The data in Figure 6 illustrates significant differences in the running time of different methods under varying user densities. The running time of the local differential privacy method gradually increases with user density, from two-time units at a user density of 5-to-19-time units at a user density of 10. The running time of the differential privacy method based on the Laplace mechanism shows a more pronounced increase, from 11-time units (user density of 5) to 160-time units (user density of 10). In contrast, the running time of the proposed method is higher than the other two methods at all user densities, rising from 14-time units at a user density of 5-to-260-time units at a user density of 10. These results indicate that as user density increases, the running time of all methods significantly rises, with the proposed method showing the most substantial increase. The experimental results demonstrate that the proposed privacy protection scheme has a higher running time compared to local differential privacy and the differential privacy method based on the Laplace mechanism under different user densities. As user density increases, the running time of all methods increases significantly due to the greater volume of user data that needs to be processed and protected, leading to increased computational complexity. Specifically, the proposed method, although exhibiting excellent privacy protection effectiveness, also shows a significant increase in running time. This is attributed to the complexity of the uncertain graph model and the hypergraph clustering method employed.

## 5 CONCLUSION

A comprehensive privacy protection scheme for mobile big data was proposed in this study, consisting of two parts: location and semantic information protection. For location information protection, an uncertain graph model was employed, jointly protecting the user layer and the location layer, effectively resisting combined attacks, and enhancing the security of location privacy. For semantic information protection, a hypergraph clustering method was used to structurally protect the user layer and the semantic layer, improving the privacy security of semantic information. The experimental results demonstrate that the proposed method shows significant performance in data usability, metric values, edge betweenness centrality values under different privacy budget thresholds, privacy protection effectiveness, and running time across different user densities. Specifically, as the privacy budget and user density increase, the edge betweenness centrality values and running time of the proposed method increase significantly. However, its privacy protection effectiveness outperforms existing methods under all tested conditions.

This study holds substantial academic value and practical significance. Firstly, by introducing the uncertain graph model and the hypergraph clustering method, effective protection of both location and semantic information was achieved, enhancing the comprehensiveness and depth of mobile big data privacy protection. Secondly, the experimental results validated the good privacy protection effectiveness and data usability of the proposed method under various conditions. However, the significant increase in running time under high user density and large privacy budget conditions indicates certain limitations of the proposed method regarding computational resource requirements. Future study can focus on the following directions: the algorithm could be optimized to reduce computational complexity and improve operational efficiency; privacy protection needs in different scenarios could be further investigated to improve protection strategies; and multidimensional data protection methods could be explored to enhance overall privacy protection

capabilities. These improvements could further increase the practicality and adaptability of the proposed method, providing robust support for the development of mobile big data privacy protection.

## 6 REFERENCES


[1] S. Anawar, N. F. Othman, S. R. Selamat, Z. Ayop, N. Harum, and F. Abdul Rahim, "Security and privacy challenges of big data adoption: A qualitative study in telecommunication industry," *International Journal of Interactive Mobile Technologies (iJIM),* vol. 16, no. 19, pp. 81–97, 2022. https://doi.org/10.3991/ijim.v16i19.32093

[2] A. Chennamaneni and B. Gupta, "The privacy protection behaviours of the mobile app users: Exploring the role of neuroticism and protection motivation theory," *Behaviour & Information Technology,* vol. 42, no. 12, pp. 2011–2029, 2023. https://doi.org/10.1080/0144929X.2022.2106307

[3] R. S. Almogbel and A. A. Alkhalifah, "User behavior in social networks toward privacy and trust: Literature review," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 16, no. 1, pp. 38–51, 2022. https://doi.org/10.3991/ijim.v16i01.27763

[4] T. Bikku, N. S. Biyyapu, J. C. Sekhar, M. K. Kumar, S. M. Nokerov, and V. K. Pratap, "The social network dilemma: Safeguarding privacy and security in an online community," *International Journal of Safety and Security Engineering,* vol. 14, no. 1, pp. 125–133, 2024. https://doi.org/10.18280/ijsse.140112

[5] R. Mueller, "Big data, big gap: Working towards a HIPAA framework that covers big data," *Indiana Law Journal,* vol. 97, no. 4, p. 10, 2022.

[6] O. Senturk and A. Baghırov, "Enhancing sustainable development through blockchain: A study on risk management and data integrity," *Journal of Organizations, Technology and Entrepreneurship,* vol. 1, no. 2, pp. 110–126, 2023. https://doi.org/10.56578/jote010204

[7] N. Wiedemann, K. Janowicz, M. Raubal, and O. Kounadi, "Where you go is who you are: A study on machine learning based semantic privacy attacks," *Journal of Big Data,* vol. 11, no. 1, p. 39, 2024. https://doi.org/10.1186/s40537-024-00888-8

[8] L. Carmi, M. Zohar, and G. M. Riva, "The European general data protection regulation (GDPR) in mHealth: Theoretical and practical aspects for practitioners' use," *Medicine, Science and the Law*, vol. 63, no. 1, pp. 61–68, 2023. https://doi.org/10.1177/00258024221118411

[9] M. D. Sajid and S. Kavitha, "Privacy-preserving photo sharing on online social networks: A review," *International Journal of Safety and Security Engineering,* vol. 14, no. 1, pp. 297–308, 2024. https://doi.org/10.18280/ijsse.140129

[10] S. Ribeiro-Navarrete, J. R. Saura, and D. Palacios-Marqués, "Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy," *Technological Forecasting and Social Change,* vol. 167, p. 120681, 2021. https://doi.org/10.1016/j.techfore.2021.120681

[11] T. N. Cooke, "Metadata, jailbreaking, and the cybernetic governmentality of IOS: Or, the need to distinguish digital privacy from digital privacy," *Surveillance & Society,* vol. 18, no. 1, pp. 90–103, 2020. https://doi.org/10.24908/ss.v18i1.13118

[12] K. R. Rao and S. Naganjaneyulu, "Permissioned healthcare blockchain system for securing the EHRs with privacy preservation," *Ingénierie des Systèmes d'Information,* vol. 26, no. 4, pp. 393–402, 2021. https://doi.org/10.18280/isi.260407

[13] Z. Iftikhar, A. Anjum, A. Khan, M. A. Shah, and G. Jeon, "Privacy preservation in the internet of vehicles using local differential privacy and IOTA ledger," *Cluster Computing,* vol. 26, no. 6, pp. 3361–3377, 2023. https://doi.org/10.1007/s10586-023-04002-0

[14] S. A. Kahate and A. D. Raut, "Comprehensive analysis of privacy attacks in online social network: Security issues and challenges," *International Journal of Safety and Security Engineering*, vol. 12, no. 4, pp. 507–518, 2022. https://doi.org/10.18280/ijsse.120412

[15] M. Giacalone, D. C. Sinitò, M. V. Calciano, and V. Santarcangelo, "A novel big data approach for record and represent compliance in the Covid-19 era," *Big Data Research,* vol. 27, no. 28, p. 100290, 2022. https://doi.org/10.1016/j.bdr.2021.100290

[16] J. Gilbert, O. Adekanmbi, and C. Harrison, "Using mobile big data to support emergency preparedness and address economically vulnerable communities during the COVID-19 pandemic in Nigeria," *Data & Policy,* vol. 3, p. e21, 2021. https://doi.org/10.1017/dap.2021.12

[17] S. Sampaio, P. R. Sousa, C. Martins, A. Ferreira, L. Antunes, and R. Cruz-Correia, "Collecting, processing and secondary using personal and (pseudo) anonymized data in smart cities," *Applied Sciences,* vol. 13, no. 6, p. 3830, 2023. https://doi.org/10.3390/app13063830

[18] P. Sanchol and S. Fugkeaw, "A fully outsourced attribute-based signcryption scheme supporting privacy-preserving policy update in mobile cloud computing," *IEEE Access,* vol. 11, pp. 145915–145930, 2023. https://doi.org/10.1109/ACCESS.2023.3341095

[19] G. U. Srikanth and L. C. Jaffrin, "Security issues in cloud and mobile cloud: A comprehensive survey," *Information Security Journal: A Global Perspective,* vol. 31, no. 6, pp. 686–710, 2022. https://doi.org/10.1080/19393555.2022.2035470

[20] M. Elhoseny, M. Siraj, K. Haseeb, M. Nawaz, M. Altamimi, and M. I. Alghamdi, "Energy-efficient mobile agent protocol for secure IoT sustainable applications," *Sustainability,* vol. 14, no. 14, p. 8960, 2022. https://doi.org/10.3390/su14148960

## 7 AUTHOR

**Peihua Su** graduated with a master's degree from Xi'an Shiyou University and is currently employed at Xi'an International University. Her primary research interests include computer applications and data analysis (E-mail: sphvsyj@163.com).