Personalized Privacy Protection Method for Sensitive Temporal Data

Ji Lei*, Wang Xinyi, Zhang Zhimin
School of Computer Science and Information Engineering, Harbin Normal University, Harbin, China
**Corresponding Author*

Abstract: Crowd sensing, as a new paradigm for perception data collection, has been widely applied. However, in the process of constructing a knowledge graph for group intelligence perception networks, attackers can use the correlation of participant temporal data to reconstruct participant related information, resulting in privacy leakage. A participant sensitive temporal data privacy protection method(PSTDPP) based on localized differential privacy is proposed to address this issue. Firstly, we use a knowledge graph to construct a sensitive temporal data relationship network and explore the correlation between temporal data background knowledge. Secondly, by combining attention mechanisms and mutual information ideas, a privacy model for participant submitted data and an availability model for perceived data are constructed, and personalized local differential privacy is used to solve the problem of graph data availability under privacy budget constraints. The simulation results show that the proposed method can effectively prevent the privacy leakage problem of sensitive temporal data, and can achieve a good balance between privacy protection level and data availability.

Keywords: Crowd Sensing; Knowledge Graph; Localized Differential Privacy; Personalized Privacy Protection

1. Introduction

As a new perception data acquisition paradigm in the field of multi-sensing devices iot [1], Crowd Sensing combines mobile sensing and crowdsourcing ideas to crowdsource tasks through mobile devices and quickly collect and acquire massive data.MCS has been widely used in such fields as transportation planning [2], environmental monitoring [3], and intelligent medical care [4].A typical MCS architecture consists of a task requester who requests the perception platform to collect data, a participant who collects the perception data via a mobile device, and a tripartite perception platform that aggregates and processes the perception data uploaded by the participant.

knowledge graph (KG), as a visual mapping tool for knowledge domain, has attracted much attention in semantic mining related work in many fields. By building a visual network to assist industry development,The existing mainstream general domain knowledge maps such as Freebase[5], DBpedia[6], YAGO[7], NELL[8], Wikidata[9], etc. have been constructed and can be applied to the actual environment.Knowledge graph is a directed network graph that takes entities in the real world as nodes and the relationship between nodes as edges.In the graph, each directed edge and its linked head and tail entities together form a fact triplet, indicating that the relationship between the head and tail entities is.This paper considers adding time information to the expression of knowledge structure in the field of swarm intelligence perception to ensure the timeliness of data, embedding time information into the embedding process of knowledge graph, adding time dimension information on the basis of fact triples, expressed as, where is the occurrence time of fact relations.

As the structural data of graph network, knowledge graph contains rich data information. The multi-modal characteristics of information in graph structure also greatly increase the difficulty of privacy definition, such as the absence or existence of nodes in the graph [11] and the absence or existence of specific edges between them [12].Differential privacy is an effective tool for data analysis and data privacy protection. As a statistical concept of privacy, differential privacy has several ideal properties: (1) It is robust to the edge information, because learning additional information about the data

generating entity will not weaken the privacy performance [13];(2) Independent of post-processing, arbitrary post-calculation of private data will not weaken privacy performance [14].

In view of the above, this paper considers the privacy of graph data while protecting the preservation of graph structure. By dividing different privacy rights into multiple edges, noise disturbance is carried out to provide sufficient privacy protection and ensure the data availability of knowledge graph.The main contributions of the proposed algorithm are as follows:

(1) Participants model the submitted data locally according to the knowledge graph construction algorithm, and incorporate the idea of localized differential privacy for noise perturbation, without relying on a trusted third party.

(2) Combined with LDP privacy budget and mutual information idea, the differential privacy protection of attribute association is realized.Avoid situations where large-scale statistical calculations on a single attribute can restore the distribution of perceived data.And the knowledge graph of sensitive time series data submitted by participants can satisfy the availability of graph data and the privacy.

2. Related Work

In recent years, researchers have made a lot of achievements in differential privacy of graph-structured data.Day et al. [16] proposed an "edge-adding" projection algorithm.Edge traversal is completed by setting all edge sets on the graph model to empty and ordering them according to the stability of the edges.At this time, if the degree between different nodes with the same link edge is less than the threshold of the global degree, the corresponding edge is added.Otherwise, the addition of this edge is excluded.However, due to the uncertainty in the order of adding edges, the algorithm cannot fully retain the edge information in the original graph structure.Liu et al. [17] proposed a local differential privacy model DP-LUSN for owning social network architecture.The DP-LUSN model first uses the community discovery to mine the potential relationship organization, then reconstructs the community structure according to the probability ordering of the edges, and instill the noise satisfying the Laplacian distribution into the community. However, the instability of the algorithm leads to a large disturbance to the original graph network structure, which may affect the availability of the graph network data.Liu et al. [18] proposed a DP-ERGM algorithm based on Random indicator graph model (EGRM) to generate node networks while ensuring differential privacy of nodes.Compared with other methods, this algorithm pays more attention to protecting the structural features of the original network model by summarizing important associations in different graph network models and inferring statistics accordingly. However, the scope of application of this algorithm is limited, and it is difficult to be widely applied to the privacy protection of various types and scales of graph networks.Hu et al. [19] designed a UGDP algorithm, which aims to map the become weights to probability values to ensure that the weight information of the edge is not leaked.This algorithm uses the method of uncertain graph to publish graph network data, but this method will lead to the loss of some structural information and ignore the availability of graph structure data.

Different from the existing methods, this paper considers the characteristics of the swarm intelligence perception environment, and adds noise to the data feature vector containing sensitive information in the perception data set by the participant perception terminal to protect the individual privacy protection needs of the participant, while avoiding the noise gradient accumulation phenomenon caused by iterative addition of noise in the model training process.Moreover, the decentralized training method can ensure the attack speculation defense against semi-trusted third-party platforms and malicious attackers, so that it can not effectively reverse infer the sensitive data of participants.

3. Personalized Privacy Protection Method for Sensitive Time Series Data

This section mainly introduces the system framework design of PSTDPP, which is mainly divided into two parts, the knowledge graph module and the privacy protection module.

3.1 System Framework

The framework of knowledge graph of sensitive time series data is shown in Figure 3. The main modules can be divided into two parts: knowledge graph learning module and privacy protection module.The knowledge graph construction module is constructed in

combination with the classical TransE model, and entity embedding information is obtained by extracting data feature vectors through features. Then, the attention mechanism is combined to build a noise-added model for the sensitive data of participants, and the feature vectors of the noise-added participants are obtained.

In the building block of knowledge graph, considering the characteristics of swarm intelligence perception data, the concept of decentralization is adopted to collect, process and model perception data locally.After privacy protection, the model satisfying the localized differential privacy is transferred to the center of the perception platform for model fusion, which ensures the accuracy of the model and satisfies the privacy protection of the participants' personalized sensitive data.

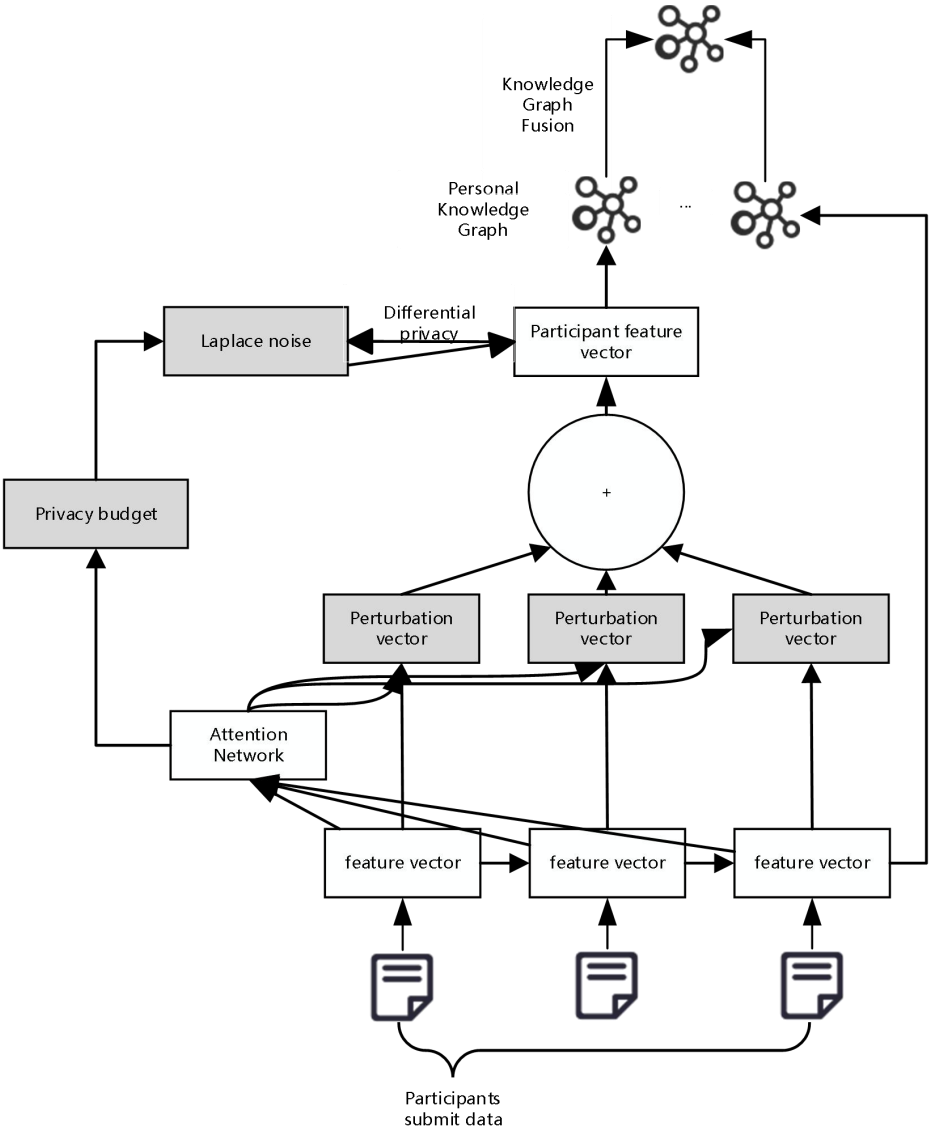

Figure 1. System Framework

3.2 Building Blocks of Knowledge Graph

Knowledge graph building module usually adopts knowledge graph representation learning in the pre-processing stage. In this paper, the classical knowledge graph representation learning method TransE is combined to carry out feature vector learning of entity attributes and relationship associations in the group intelligence perception environment.Although most current knowledge graph representation learning methods can mine and analyze graph network structure information in related fields, this paper considers the characteristics of swarm intelligence perception data set and combines the idea of decentralization to build personal knowledge graph through local training of participants. After privacy protection, the

perception platform is integrated and summarized to ensure the accuracy of knowledge graph model.Table 1 describes the complete steps of the TransE algorithm.

Table 1. TransE Knowledge Graph Embedding Algorithm

Algorithm 1: Knowledge graph training algorithm
Input: Training set $= \{(e_i, r, e_j)\}$,Entity set E and relation set R,margin value γ,Embedded vector dimensions k
1. Initialize for each relationship vector $r \in R$,Random sampling from the $(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ interval.
2. For each relation vector $r \in R$,Divide by its own L2 norm.
3. For each entity vector$e_* \in E$,Random sampling from the $(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ interval
4. circulate:
5. For each entity vector $e_* \in E$,Divide by its own L2 norm.
6. Remove the samples with a number of b from the training set S as one S_{batch}.
7. Initialize the triplet set T_{batch} as an empty list
8. ergodic:$(e_i, r, e_j) \in S_{batch}$,carry out:
9. Replace the head or tail entity of the correct triples (e_i', r, e_j) or (e_i, r, e_j').
10. Both the positive and negative sample triples are placed in the T_{batch} list.
11. The traversal ends.
12. Update the entity, relationship vectors according to the gradient descent.
13. loop end.

3.3 Privacy Protection Module

In the privacy protection module, in order to protect participants' perceived data privacy, this paper adds Laplacian noise to the correlation vector of the constructed knowledge graph.Considering the utility of the data, first of all, different levels of noise are added to the privacy data of the participants with different sensitivity levels.The data density submitted by participants reflects different preference characteristics of participants, and the weight obtained based on the attention mechanism reflects the participants' concern about different entity relationship data. The higher the weight ratio, the greater the impact of the perceived data submitted by participants on the information privacy of participants.

The budget allocation system combining the attention mechanism with the noise addition mechanism can ensure the utility of the data and the privacy protection of the participant data in the process of adding noise to the participant feature vector.The specific algorithm flow is shown in algorithm 2. The formula for calculating the global sensitivity of the feature vector is as follows: 3.3 Privacy protection module

In the privacy protection module, in order to protect participants' perceived data privacy, this paper adds Laplacian noise to the correlation vector of the constructed knowledge graph.Considering the utility of the data, first of all, different levels of noise are added to the privacy data of the participants with different sensitivity levels.The data density submitted by participants reflects different preference characteristics of participants, and the weight obtained based on the attention mechanism reflects the participants' concern about different entity relationship data. The higher the weight ratio, the greater the impact of the perceived data submitted by participants on the information privacy of participants.

The budget allocation system combining the attention mechanism with the noise addition mechanism can ensure the utility of the data and the privacy protection of the participant data in the process of adding noise to the participant feature vector.The specific algorithm flow is shown in algorithm 2. The formula for calculating the global sensitivity of the feature vector is as follows:

$$\Delta f_k = \max\{e(c_k^i)\} - \min\{e(c_k^i)\} \#(3)$$

Where it represents entities in the knowledge graph and represents perceptual data submitted by participants.

Then, these feature vectors are assigned with different levels of privacy budget combined with sensitivity weight, and the privacy budget is assigned to the feature vector with greater sensitivity weight and less noise disturbance.In this regard, the privacy budget formula for eigenvector allocation is as follows:

$$\varepsilon_k = s_{c_k^i, c_j} * \varepsilon_1 \#(4)$$

In order to ensure the privacy protection of participants' sensitive time series data during the construction of knowledge graph, noise is further added to the knowledge graph of submitted perception data after the privacy protection of personal knowledge graph is uploaded. The feature vector of participants after adding noise is shown as follows:

$$\widehat{e(i)} = \sum_{k=1}^{N_i} \widehat{e(c_k^i)} + \text{Lap}\left(\frac{\Delta f}{\varepsilon_2}\right) \#(5)$$

Table 2. Knowledge Graph Privacy Protection Algorithm

Algorithm 2: Privacy protection algorithm
Input: Knowledge graph G, privacy budget ε; Output: Knowledge graph that satisfies differential privacy $\overline{G}$; 1: Initialize the hyperparameters. 2:Extract the participant feature vector from the knowledge graph, learn to obtain the embedding of entities and the embedding of relationships, and learn the participant information to obtain the corresponding vector. 3:By the attention mechanism, get the participants to submit the perceptual data c_k^i For the current map c_j Impact weight of the construction $s_{c_k^i, c_j}$. 4:A noise is added to the feature representation of the map to calculate the global sensitivity: $\Delta f_k = \max\{e(c_k^i)\} - \min\{e(c_k^i)\}$. 5: for(int $m = 0; m < k; m++$){ 6: $\varepsilon_k = s_{c_k^i, c_j} * \varepsilon_1$; 7: $\widehat{e(c_k^i)} == e(c_k^i) + \text{Lap}\left(\frac{\Delta f_k}{\varepsilon_k}\right)$;} 8:Add noise to the user feature vector $\widehat{e(i)} = \sum_{k=1}^{N_i} \widehat{e(c_k^i)} + \text{Lap}\left(\frac{\Delta f}{\varepsilon_2}\right)$; 9:Add the user noise feature vector fusion into the knowledge graph.

Theorem 1 PSTDPP Algorithmic satisfaction (ε_1+ε_2)- Differential privacy.

Prove

In Algorithm 2, the privacy protection module is divided into two parts. In the first part, the sensitivity of the feature vector of the perceptual data is calculated, and the corresponding global sensitivity is shown in the following formula:

$$\Delta f_k \max\{e(c_k^i)\} \min\{e(c_k^i)\} = 1 \#(6)$$

Where the eigenvector $e(c_k^i)$ is a constant vector, where $k = 1,2, \ldots, N_i$.

The overall privacy of this stage is counted as ε_1 ,The allocated privacy budget is $\varepsilon_k = \varepsilon_k = s_{c_k^i, c_j} * \varepsilon_1$,,Therefore, this stage is satisfied ε_1 -Differential privacy;In the second part, the participants perceive the data feature vectors e(i)The corresponding global sensitivity is given by 1,Assign a privacy budget of that is ε_2,Noise addition satisfaction $\text{Lap}(1/\varepsilon_2)$ distribute noise, so the noise is satisfied ε_2 -Differential privacy.The above two parts correspond in the algorithm to noise additions in lines 7 and 8, respectively, and the perturbed objects are both feature vectors of participant perception data, combined with definitions 4,This algorithm satisfies ε- Differential privacy, among $\varepsilon = \varepsilon_1 + \varepsilon_2$

4. Simulation Experiment

4.1 Environment Settings

The simulation environment is Win11 operating system, the graphics card is configured as NVDIA Force RTX3050TI Laptop GPU 4GB, the processor is configured as AMD Ryzen 7 5800H, and the compilation environment is python 3.9.18.In addition, the data set in this paper has been extracted by knowledge, and the data format is shown in Table 3:

Table 3. Data Format

Head entity	Tail entity	relation
UID	Sensor_type	Sensor_value
UID	Sensor_device_type	Sensor_sampling_time

4.2 Privacy Analysis

In order to study the impact of privacy budget and segmentation parameters on privacy, the virtual data was randomly generated to evaluate the privacy of the PSTDPP algorithm proposed in this paper.

(1) Match degree of attack results

By taking the published graph after privacy disturbance as a subgraph and taking it as the knowledge background, we observe the match

degree between the attack results obtained after the subgraph attack and the original real data set. When the match degree is high, the privacy of the model is worse; on the contrary, the privacy of the model is better.At the same time, it is assumed that the malicious attacker grasps the background knowledge and incomplete correlation subgraph of the participant, and then carries out node attack on the knowledge graph publishing graph.The experimental process is the change of match degree of attack results with the increase of segmentation coefficient k and the allocation of different privacy budgets.As shown in Figure 5.

As can be seen from FIG. 5:1) As the number of subgraphs decreases, that is, the segmentation coefficient increases, the match degree of attack results shows a downward trend, and when k= about 40, the match degree reaches the minimum value, and the privacy protection effect is the best.2) In the differential privacy model, the smaller the privacy budget, the higher the degree of privacy protection, the better the effect of privacy protection, and the lower the matching degree of attack results.3) When the segmentation coefficient is too high, the subgraph contains significantly less data, and the match degree of attack results will rebound and increase, and the privacy protection effect will gradually deteriorate.

(2) Privacy protection

The experiment compares the privacy protection level of the graph after perturbation by PSTDPP and differential privacy algorithm under different privacy budgets, as shown in Figure 6.

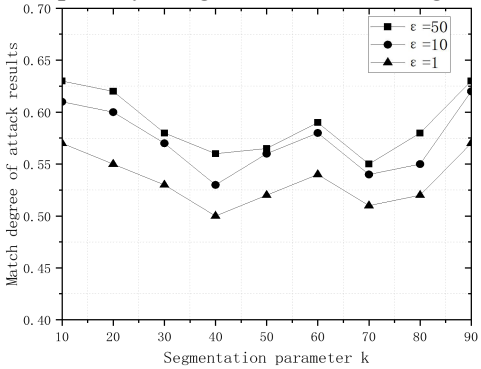

Figure 2. The Matching Degree of Attack Results of PSTDPP Algorithm

It can be seen from Figure 6 that under different privacy budgets, compared with the DP scheme that directly performs differential privacy on the graph, the proposed PSTDPP scheme in this paper shows better privacy protection effect, which proves the theoretical support in this paper.In other words, the privacy protection of graph data needs to consider the edge structure and the reasonable allocation of privacy budget.With the increase of privacy budget, when the increase to a certain value, because the amount of noise introduced is small, the privacy protection of the two methods is less different.

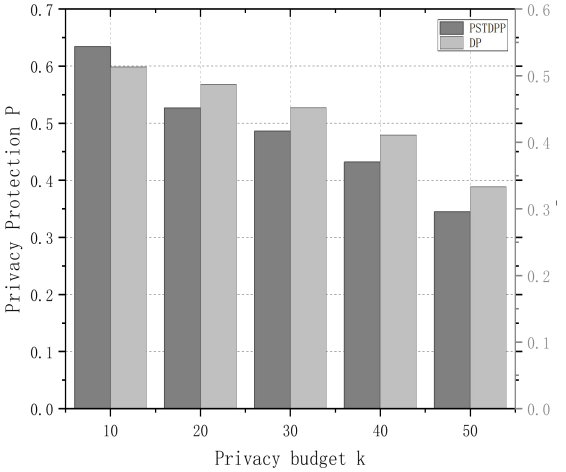

Figure 3. The Trend of Changes in Privacy Protection Level of Pstdpp Algorithm

5. Conclusion

In this paper, the construction process of knowledge graph is integrated into the swarm intelligence awareness environment, and the relational association mining is carried out on the swarm intelligence awareness data set. In view of the sensitive data leakage and privacy risks concerned by participants during the construction process of knowledge graph, the differential privacy model with good robustness against background knowledge attacks is integrated into the knowledge graph embedding process.Therefore, a knowledge graph embedding method of swarm intelligence sensing sensitive temporal data with differential privacy is proposed.Considering the different performance scenarios of terminal devices in swarm intelligence environment, this method integrates differential privacy into the construction process of knowledge graph model to reduce the cost and ensure the privacy of participants.This method is suitable for swarm intelligence sensing multi-terminal scenario, and the experiment verifies that it can ensure the effectiveness of knowledge graph modeling while protecting data privacy and security.Future work will further consider how to optimize the knowledge graph construction algorithm to reduce the computational overhead while providing privacy protection, so as to

achieve a balance between model accuracy, algorithm performance and privacy protection.

References

[1]Ganti R K, Ye F, Lei H. Mobile crowdsensing: current state and future challenges[J]. IEEE communications Magazine, 2011, 49(11): 32-39.

[2]Ali A, Qureshi M A, Shiraz M, et al. Mobile crowd sensing based dynamic traffic efficiency framework for urban traffic congestion control[J]. Sustainable Computing: Informatics and Systems, 2021, 32(100608): 1-8.

[3]Alvear O, Calafate C T, Cano J C, et al. Crowdsensing in smart cities: Overview, platforms, and environment sensing issues[J]. Sensors, 2018, 18(2): 460-488.

[4]Pryss R, Reichert M, Langguth B, et al. Mobile crowd sensing services for tinnitus assessment, therapy, and research[C]//2015 IEEE International Conference on Mobile Services. Florence, Italy, IEEE, 2015: 352-359.

[5]Auer S, Bizer C, Kobilarov G, et al. Dbpedia: A nucleus for a web of open data[C]//international semantic web conference. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007: 722-735.

[6]Suchanek F M , Kasneci G , Weikum G .YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia[C]//International Conference on World Wide Web. 2007: 697–706.

[7]Carlson A, Betteridge J, Kisiel B, et al. Toward an architecture for never-ending language learning[C]//Proceedings of the AAAI conference on artificial intelligence. Westin Peachtree Plaza, Atlanta, GA, USA. 2010, 24(1): 1306-1313.

[8]Vrandečić D, Krötzsch M. Wikidata: a free collaborative knowledgebase[J]. Communications of the ACM, 2014, 57(10): 78-85.

[9]Berant J, Chou A, Frostig R, et al. Semantic parsing on freebase from question-answer pairs[C]//Proceedings of the 2013 conference on empirical methods in natural language processing. Seattle, Washington, USA. 2013: 1533-1544.

[10]Kasiviswanathan S P, Nissim K, Raskhodnikova S, et al. Analyzing graphs with node differential privacy[C]//Theory of Cryptography: 10th Theory of Cryptography Conference, TCC 2013, Tokyo, Japan, March 3-6, 2013. Proceedings. Springer Berlin Heidelberg, 2013: 457-476.

[11]Karwa V, Raskhodnikova S, Smith A, et al. Private analysis of graph structure[J]. Proceedings of the VLDB Endowment, 2011, 4(11): 1146-1157.

[12]Kasiviswanathan S P, Smith A. On the'semantics' of differential privacy: A bayesian formulation[J]. Journal of Privacy and Confidentiality, 2014, 6(1): 1-16.

[13]Dwork C ,McSherry F ,Nissim K , et al.Calibrating Noise to Sensitivity in Private Data Analysis[J]. Journal of Privacy and Confidentiality, 2017, 7(3): 17-51.

[14]Dwork C. Differential privacy: A survey of results[C]//International conference on theory and applications of models of computation. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008: 1-19.

[15]Dwork C, Lei J. Differential privacy and robust statistics[C]//Proceedings of the forty-first annual ACM symposium on Theory of computing. New York, NY, United States. 2009: 371-380.

[16]Day W Y, Li N, Lyu M. Publishing graph degree distribution with node differential privacy.[C]//Proceedings of the 2016 International Conference on Management of Data. San Francisco, USA,2016:123-138

[17]Liu P, Xu Y X, Jiang Q, et al. Local differential privacy for social network publishing[J]. Neurocomputing, 2020, 391: 273-279.

[18]Liu F, Eugenio E, Jin I H, et al. Differentially Private Generation of Social Networks via Exponential Random Graph Models[C]//2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). Madrid, Spain. IEEE, 2020: 1695-1700

[19]Hu J, Yan J, Wu Z Q, et al. A Privacy-Preserving Approach in Friendly-Correlations of Graph Based on Edge-Differential Privacy[J]. Journal of Information Science & Engineering, 2019, 35(4): 821-837.

