JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 37, 885-899 (2021)
DOI: 10.6688/JISE.202107_37(4).0009

A Differential Privacy Topology Scheme for Average Path Length Query*

TONG DONG, YONG ZENG, ZHI-HONG LIU, JIAN-FENG MA AND XIAO-YAN ZHU
School of Cyber Engineering
School of Telecommunications Engineering
Xidian University
Xi'an, 710071 P.R. China
E-mail: {tdong; yzeng; zhhliu; jfma; xyzhu}@mail.xidian.edu.cn

Privacy protection of sensitive information has become an urgent problem to be solved in social networks. Differential privacy is used in many privacy protection methods because it can provide strong protection. Most of existing differential privacy scheme mainly implements the privacy protection on nodes or edges in the network by perturbing the data query results. The privacy protection requirements of multiple types of information cannot be satisfied in these schemes. In order to solve these issues, a differential privacy security mechanism with average path length (APL) query is proposed in this paper, which realize the privacy protection of both edge weights and network vertices. The reasons for choosing this attribute as the query function are analyzed. The global sensitivity of APL query under the need of node privacy protection and edge-weighted privacy protection is proved. Based on previous studies, the concept of edge-weighted neighborhood graph in differential privacy is proposed. The relationship between data availability and privacy control parameters in differential privacy is analyzed through experiments.

Keywords: social network, differential privacy, network topology, privacy protection, global sensitivity

1. INTRODUCTION

With the rapid development of social networks, more and more individuals will have very complex interactions. The process of interaction between individuals in society is a typical social network model, which plays an important role in production and life. The research on privacy protection algorithms of social networks has become an indispensable topic when discussing social networks. In differential privacy, since the network topology information contains more private information, publishing the network topology related query results has also become a trend. The query for network topology information is divided into degree distribution, shortest path distribution and other overall distribution query, which leads to the release of the query function is very sensitive to the network scale.

Differential privacy was first proposed by Dwork [1, 2] and given strict definition and proof. Currently, most of differential privacy schemes only focus on the single data privacy of nodes or edges, these methods cannot meet the requirements of privacy protection in the network.

X. Xiao *et al.* [3] proposed a framework that applies wavelet transforms on the data before adding noise to it, the instantiations for both ordinal and nominal data are presented.

Received January 17, 2020; revised December 15, 2019; accepted March 14, 2020.
Communicated by Xiaohong Jiang.
* This work was sponsored by the National Natural Science Foundation of China (No. 61941105) and China 111 Project Foundation (No. B16037).

The core of their solution is a framework that applies wavelet transforms on the data before adding noise to it. This method provides a theoretical analysis on their privacy and utility guarantees. [4] proposed distributed algorithms for in-network tracking and range queries for aggregated data. This scheme stores the target detection information locally in the network and answers a query by examining the circumference of the given range. But the cost of updating data about mobile targets is proportional to the target displacement, and this may bring more cost. [5] proposed a method which published the degree distribution and cut the query and the shortest path information to maintain the privacy of the node. In this paper, it is shows that the amount of noise is significantly reduced from O(n) to O($log(n)$). Pinot *et al.* [6] proposed a differential privacy scheme based on the release of minimum spanning tree clustering, which realized the privacy protection of the weights. They present the first differentially private clustering method for arbitrary-shaped node clusters in a graph, and their algorithm is theoretically motivated. However, the sensitivity of these query functions is too large, and the privacy protection of the node is more concerned in these query functions.

Karwa *et al.*, [7, 8] proposed a perturbation based on network structure query, which can protect the edges information in the network well. Hay *et al.* [9] proposed a differential privacy algorithm based on edge privacy protection, which released the perturbation result of the moderate distribution through constraint reasoning technology. The techniques can be used for estimating the degree sequence of a graph very precisely, and for computing a histogram that can support arbitrary range queries accurately. Chen *et al.* [10] used the exponential mechanism to add the constructed noise adjacency matrix, they released the perturbed network structure information such as degree distribution, cutting query and shortest path to realize edge weight protection. This is the first work providing a practical solution for publishing real-life network data via differential privacy. Sealfon *et al.* [11] proposed a policy based on data distribution. This method only cares about edge privacy protection and has a high query cost. Brunette *et al.* [12] proposed a differential privacy based on Laplace mechanism, which distributes edge weight noise based on the noise threshold of edge subsets, and preserves the spectral information of the input pattern while guaranteeing ε differential privacy. Their mechanisms guarantee ε-differential privacy for a reasonable level of privacy ε, while preserving the spectral information of the input graph. Li *et al.* [13] treat edge weight sequences as unallocated histograms and add noise based on histograms to achieve edge-weighted differential privacy. This approach effectively improved the accuracy and utility of the released data. Lan Lihui *et al.* [14] proposed differential privacy based on WSQuery model. This scheme adds noise to the mapping vector of query results to realize the privacy protection of edge and edge weights. Wang Hong [14] proposed the DP-OPTICS algorithm by preprocessing the data and combining the existing algorithms, which improved the data utility. All of these methods have good efficiency, but they only care about edge privacy protection, besides the balance of data availability and security cannot be maintained.

In order to simplify the operation process of differential privacy and take it into account the nodes and edges privacy in the network, this paper proposes a differential privacy algorithm based on network average path length (APL) query. The contributions of this article are as follows:

- Through analysis, the network APL is selected as the target of differential privacy.

- The APL under ER and BA networks is analyzed and combined with the previous studies, the concept of edge weighted neighbor graph is given.
- The global sensitivity of satisfying the dual privacy protection requirements of node and edge weight is obtained.
- The relationship between the availability of data and the privacy control parameters is analyzed through experiments.

2. DIFFERENTIAL PRIVACY SCHEME FOR APL

The topology-based differential privacy algorithm is proposed in this paper. The purpose of this algorithm is to realize the dual privacy protection of node and edge weight by publishing the perturbed query result of network APL. This strategy is based on the same kind of differential privacy algorithm to achieve the purpose of node privacy protection and edge weight privacy protection. In real world the scale of actual networks is generally much larger than 100. In this paper, the smaller network has not been considered, *i.e.*, the minimum network size is set to 100.

2.1 Differential Privacy Sensitivity Analysis for Node Protection

In a network, APL refers to the average of the shortest path lengths between all nodes in the network. When querying the APL, the interference of the query result is used to blur whether a node exists, thereby implementing privacy protection of the node.

(A) APL analysis

According to the corresponding literature, the expressions of APL in different types of unweighted networks are obtained.

(a) APL of ER networks

The distance $L(i, j)$ between two nodes i and j in the unweighted network is defined as the number of edges connecting the two nodes. The APL is shown in Eq. (1).

$$\langle L \rangle = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i,j} L(i,j) \tag{1}$$

According to [15-17], the unweighted ER network APL can be obtained as Eq. (2):

$$\langle L \rangle_{ER} \approx \frac{\log N}{\log \tilde{k}}. \tag{2}$$

Where N is the network size, and k is the average node degree of the network.

(b) APL of BA networks

The APL of the BA network is obtained according to [18, 21] and shown in Eq. (3).

$$\langle L \rangle_{SF} \propto \frac{\log N}{\log \log N} \tag{3}$$

It can be known from Eq. (2) that the BA network APL has a proportional relationship with the log value of the network node. The proportional relationship is shown by Eq. (4).

$$\langle L \rangle_{SF} = C \frac{\log N}{\log \log N} \tag{4}$$

Where C is a constant, *i.e.*, the positive relationship is determined by a constant.

This chapter adopts the Laplace Mechanism differential privacy scheme. The confirmation of global sensitivity is the basis for the implementation of the differential privacy scheme.

(B) Sensitivity Confirmation

The APL expressions for both networks are given in section (A), which is the basis for analyzing the sensitivity in differential privacy. The sensitivity of the two representative model networks under the scheme to APL is analyzed separately.

Definition 1: (Global sensitivity [19]**)** For any query function, the global sensitivity of the function is:

$$\triangle f = \max_{(D_1, D_2)} \| f(D_1 - f(D_2) \|_1 . \tag{5}$$

Theorem 1: In the unweighted ER network with the node size of $N(N \geq 100)$ and the existence probability of p, when the node-based privacy protection, the global sensitivity of the differential privacy for the network APL query is $Clog(100/99)$, where C is a constant.

***Proof*:** It can be known from Eqs. (1) and (2) that the sensitivity in the differential privacy policy of this paper can be obtained according to the following process:

$$\left\| f(D_1) - f(D_2) \right\|_1 = \left\| \frac{\log N}{\log \tilde{k}_1} - \frac{\log(N-1)}{\log \tilde{k}_2} \right\| . \tag{6}$$

In the ER network, the average degree is related to the node size and the edge existence probability, so the two are approximately equal, Eq. (6) can be expressed as Eq. (7):

$$\begin{aligned} \left\| f(D_1) - f(D_2) \right\|_1 &\approx \left\| \frac{\log N - \log(N-1)}{\log \tilde{k}} \right\| = \left\| \frac{\log \frac{N}{N-1}}{\log \tilde{k}} \right\| \\ &= \left\| C * \log \frac{N}{N-1} \right\| = C * \left\| \log(1 + \frac{1}{N-1}) \right\| . \end{aligned} \tag{7}$$

Where C is a constant.

It can be seen from the above analysis that $\|f(D_1) - f(D_2)\|_1$ decreases with the node size N. In this paper, the smaller network has not been considered, *i.e.*, the minimum network size is set to 100, *i.e.*, $N \geq 100$. Therefore, for the unweighted ER network, the global sensitivity of the query to the APL can be expressed as shown in Eq. (8):

$$\left\| f(D_1) - f(D_2) \right\|_1 \le C \log \frac{100}{99}. \tag{8}$$

The global sensitivity Δf of the query function in the ER network is Eq. (9):

$$\Delta f = C \log \frac{100}{99}. \tag{9}$$

In summary, in the unweighted ER network, the global sensitivity of the query function for node privacy protection is Clog(100/99). The global sensitivity of unweighted BA network is analysis in follow discussion.

Theorem 2: In an unweighted BA network with a node size of $N(N \ge 100)$, when node-based privacy protection, the global privacy sensitivity of the differential privacy for network APL queries is C(loglog100), where C is a constant.

***Proof*:** The sensitivity of the BA network APL can be determined by Eqs. (1) and (3):

$$\left\| f(D_1) - f(D_2) \right\|_1 = C \left\| \frac{\log N}{\log \log N} - \frac{\log(N-1)}{\log \log(N-1)} \right\|. \tag{10}$$

For the sake of analysis, let $\log N = T$, it can be reach that:

$$\begin{aligned} \left\| f(D_1) - f(D_2) \right\|_1 &= C \left\| \frac{\log(N-1)}{\log \log(N-1)} - \frac{\log N}{\log \log N} \right\| \\ &\le C \left\| \frac{T * \log(T-1) - (T-1) * \log T}{\log T * \log(T-1)} \right\| \\ &= C \left\| \frac{(T-1) * \log(T-1) + \log(T-1) + \log(T-1)}{\log T * \log(T-1)} \right\| \\ &= C \left\| \frac{(T-1) * \log \frac{T-1}{T} + \log(T-1)}{\log T * \log(T-1)} \right\|. \end{aligned} \tag{11}$$

Since $\frac{T-1}{T} < 0$, the above analysis can be expressed as Eq. (12):

$$\left\| f(D_1) - f(D_2) \right\|_1 \le C \left\| \frac{\log(T-1)}{\log T * \log(T-1)} \right\| = C \left\| \frac{1}{\log T} \right\|. \tag{12}$$

Since $T = \log N$, $\|f(D_1) - f(D_2)\|_1$ decreases with the node size N, The smaller network is still not considered in the BA network, and the minimum network size is still set to 100, *i.e.*, $N \ge 100$. Therefore, Eq. (13) can be obtained.

$$\left\| f(D_1) - f(D_2) \right\|_1 \le \frac{C}{\log \log 100} \tag{13}$$

From the definition of the sensitivity in Eq. (8), for the BA network, the sensitivity in the differential privacy proposed in this chapter can be expressed as Eq. (14):

$$\Delta f = \frac{C}{\log\log 100}. \tag{14}$$

In summary, in the unweighted BA network, the global sensitivity of the query function for node privacy protection is $C\log(100/99)$.

Through the above analysis, the threshold responding to the global sensitivity of the differential privacy for the network APL query can be obtained. In an actual network, the network size is usually larger. The two models analyzed in this paper can represent a part of the actual network. When querying for network APL, the threshold of global sensitivity is small, the sensitivity of the analysis is the threshold in extreme cases [22].

Therefore, when the global sensitivity of the query function APL is set to 1, the requirement of node privacy protection in the unweighted network can be met.

Theorem 3: In a weighted social network with a node size of N, the global privacy sensitivity of the differential privacy for the network APL query is 1 when the node performs privacy protection.

***Proof*:** The privacy protection of the network is also suitable for the weighted network, it is also necessary to protect the individual network. Network APL relies on the shortest path length between nodes. In a weighted network, the length of the shortest path depends not only on the number of connections between network nodes, but also on the weights. Assuming that the largest weight in a network is w, the relationship between the shortest path length L_1 in the weighted network and the shortest path length L_2 in the unweighted network can be expressed as $L_1 \le wL_2$. APL is the average of all shortest path lengths, so the APL of the weighted network will be affected by the weighting factor w compared to the APL of the unweighted network. From Eqs. (13) and (14), it can be known the global sensitivity of the unweighted network multiplied by the weighting factor w in the network is the global sensitivity of the corresponding weighted network.

The global sensitivity of a weighted ER network is shown as Eq. (15).

$$\Delta f = Cw\log\frac{100}{99} \tag{15}$$

Eq. (16) represents the global sensitivity of a weighted BA network.

$$\Delta f = \frac{Cw}{\log\log 100} \tag{16}$$

The weight of the network may be large, and the global sensitivity is small. The weight of the network is usually normalized to a certain range. In order to maintain the consistency of the node privacy protection and the superiority of APL, the weight of the network is normalized to the range of 0-1. When the privacy sensitivity of the node is set to 1 for node protection, the differential privacy requirement can still be satisfied.

2.2 Differential Privacy Sensitivity Analysis for Edge Weight Protection

The global sensitivity of query functions for node privacy protection in differential privacy is studied in Section 2.1. However, the weights in social networks are also sensitive information. In the existing differential privacy scheme, the edge weight distribution is used as the query function, and noise is added to the distribution to protect weight information. But only the overall characteristics of the weights are guaranteed, and privacy protection for specific edge weights is not involved. Therefore, a differential privacy scheme based on individual edge weight privacy protection is proposed.

(A) Edge weight differential privacy concept

The purpose of privacy protection is that even if an edge weight changes only by 1, the output of the query function cannot be distinguished by probability. A new neighbor graph concept is defined and privacy of edge weight is concerned in this section.

Definition 2: (Edge weight neighbor graph) For any one of the graph G_1, the edge weight neighbor graph G_2 is a graph having a side weight difference of 1 from the graph G_1.

(B) APL analysis and sensitivity confirmation

According to Eq. (4), APL is related to the shortest path length between all pairs of nodes in the network. The APL of the network is represented by L. Graphs G_1 and G_2 are defined as neighbor graphs and only one edge $e_{i,j}$ has different weights. The weights of the edge are expressed as w_1 and w_2 in the two graphs.

Theorem 4: In a weight social network with a node size of N, when weight-based privacy protection is performed, the global sensitivity of differential privacy is 1 for network APL queries.

***Proof*:** It is assumed that the number of shortest path in the network is K, where a shortest path P between two nodes i and j passing through an edge e. Since the weight of the edge in the neighbor graph is only 1 different from the weight of the original graph (assuming the weight is increased by 1 compared with the original graph), if the shortest path is unchanged in the neighbor graph, the length of the shortest path is increased 1. If the shortest path P changes, it does not pass the edge, it means that the shortest path length in the neighbor graph is increased by 1 compared with the original shortest path length. So, when the weight of one side of the neighbor graph is increased by 1, the variation of the shortest path length is 1 regardless of whether the shortest path between the pair of nodes changes or not. When the weight of this edge is decreased by 1, the variation of the shortest path length between the pair of nodes is also 1.

So, for a certain edge e in the original graph, when the weight w is increased by 1 or decreased by 1, the length of a shortest path passing through the edge must be increased by 1 or decreased by 1. In the worst case, the length of the K shortest paths in the neighbor graph changes, the total amount of path length of the K shortest paths change is K. The K shortest paths are represented by $P_1, P_2, \ldots, P_K$, the corresponding paths in the corresponding neighbor graphs are represented by $P'_1, P_2, \ldots, P'_K$. The lengths of the original graph and

the neighbor graph are expressed as l_1, l_2, …, l_K and l'_1, l'_2, …, l'_K, the above analysis can be expressed as:

$$\|l_i - l'_i\| \le 1, \left\|\sum_K l_i - \sum_K l'_i\right\| \le K, \left\|\frac{\sum_K l_i}{K} - \frac{\sum_K l'_i}{K}\right\| \le 1. \tag{17}$$

According to Eq. (1), the APL threshold of the neighbor graph is written as follows:

$$\|f(D_1) - f(D_2)\| \tag{18}$$

In other words, under the edge weight differential privacy goal, the sensitivity is 1 when the Laplace mechanism is implemented in the APL query. It can be written as:

$$\Delta f = 1. \tag{19}$$

Therefore, the global sensitivity of the query function analyzed in this paper is 1 for edge weight privacy protection. The threshold of shortest path variation is obtained in the worst case, this threshold always meets the actual demand in practical applications.

Algorithm 1: Differential Privacy Algorithm for Topology Query

Input: Original Graph G, Query function f, Privacy Control Parameter ε
Output: $F(G)$

1: Get Δf according to input G
2: Set $\mu = 0$ and $b = \Delta f/\varepsilon$
3: Get noise $N \sim Y(\mu, b)$
4: $f(G) \leftarrow$ Average Path Length(G)
5: $f(G) \leftarrow f(G) + N$
6: **return** $F(G)$

2.3 Description of Algorithm

In this section, the global sensitivity of the privacy protection scheme proposed in this paper is analyzed when the node or edge weight privacy protection is concerned.

According to the analysis of the global sensitivity, it is known that setting the global sensitivity to 1 for the node and edge weight privacy protection fully satisfies the requirements of differential privacy. The appropriate query function f can be obtained through the above analysis in subsection A and B. The implementation process of APL algorithm is shown in Algorithm 1.

Firstly, according to the input network and the proof of the algorithm part of this paper, the threshold Δf of differential privacy can be derived by original graph G.

Secondly, the perturbed Laplace noise can be obtained by default privacy control parameter ε. The perturbed graph G' is obtained by adding the Laplace noise calculated in the previous step to the average path of the original graph. G' is the new network obtained by executing the differential sensitivity algorithm on G.

3. EXPERIMENT AND RESULT ANALYSIS

3.1 Experimental Environment and Data

The data used in this paper comes from the model network and the real data set. The real data sets are user trust weighted directed network from Stanford Large Network Dataset Collection. In these networks, the weight range is −10 (completely untrusted) to 10 (full trust). If there is no transaction between the two users, the trust relationship value between the two users is set to 0. The specific information of the two networks is shown in Table 1.

Table 1. Network basic statistics.

Data Sets	Nodes	Edges	Weight range
soc-sign-bitcoin-otc	3,783	24,186	[−10, 10]
soc-sign-bitcoin-alpha	5,881	35,592	[−10, 10]
p2p-Gnutella05	8846	31839	[−10, 10]
p2p-Gnutella04	10876	39994	[−10, 10]
p2p1	6301	20777	[−10, 10]

An APL query is proposed in this paper to deal with the privacy protection problems of user presence privacy and trust between users on two bitcoin trading platforms. The evaluation of data availability needs to compare the original query results with the query results after the perturbation. In this paper, the query result APL is a single value, so the result of a query is prone to contingency. In order to avoid this phenomenon and make the measurement universal, the network is queried multiple times. In the two contrast vectors, one consists of the APL of the original network, the other consists of the perturbed APL. The data availability is compared by measuring the similarity between the vectors formed by the raw data and the vectors formed by the perturbed data. The higher similarity, the more common information available. The Euclidean distance [19] and the cosine distance are selected.

3.2 Analysis of Disturbance

The privacy control parameter ε in the differential privacy can measure the ability of the random algorithm F to resist attacks. When ε is small, according to the definition of differential privacy [1], the probability of judging from the original data set or from the neighbor data set is smaller, the protection is relatively greater [18]. Because the Laplace mechanism is adopted, the noise added in the perturbation of the query result obeys the Laplacian distribution. Scale parameter b is related to global sensitivity and ε. b is inversely proportional to ε. When global sensitivity is determined, the smaller ε is, the larger b is. The added noise is greater, ability to withstand attacks and availability of data is reduced. The impact of ε values on data availability is explored in experiment.

A differential privacy policy is implemented on two actual networks to observe the difference between the APL after using algorithm and the original network APL. Under each ε value, 20 correlation experiments are performed on the two networks.

As can be seen from Figs. 1 and 2, the perturbations caused by different privacy parameter values ε are quite different. The ε in the experiment are 0.1, 0.4, 0.7 and 1. As

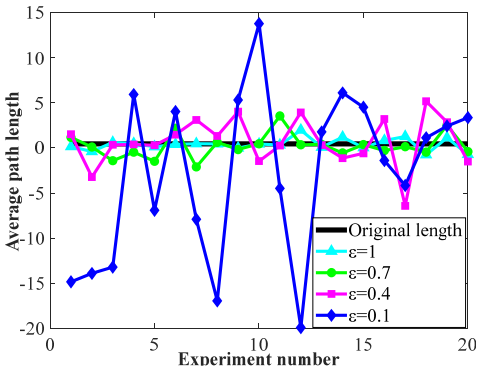

Fig. 1. Comparison of the soc-sign-bitcoin-otc data set before and after privacy protection.

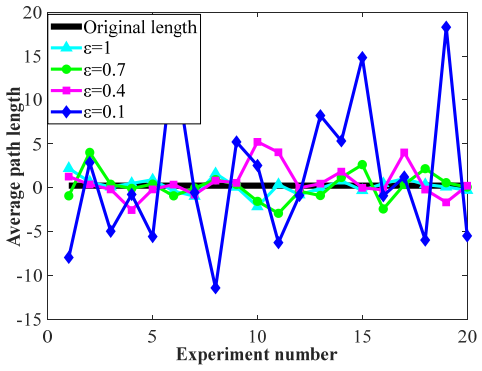

Fig. 2. Comparison of the soc-sign-bitcoin-alpha data set before and after protection.

can be seen from Fig. 2, as the ε increases, the corresponding disturbance length curve is closer to the original length. When $\varepsilon = 0.1$, the difference between the disturbance curve and the original curve is the largest. When $\varepsilon = 1$, the disturbance curve is closest to the original curve. This experiment generally shows the impact of the privacy parameter ε on the perturbation.

3.3 Analysis of Differential Privacy Result for APL Query

In the above section, two evaluation indicators are selected to evaluate the effect of differential privacy. The valid data in this experiment is a 1×20 vector. Each vector is the average of 50 experimental results. Each result of 50 experimental is from the distance (Euclidean distance and Cosine distance) between the two 1×20 vectors X and Y.

(A) Euclidean distance

The Euclidean distances of the ER and BA network under different privacy control parameters ε are shown in Fig. 3. The network scale is 2000. The scale parameter of the Laplacian distribution is $b = \Delta f/\varepsilon$. It can be seen from Fig. 3 that when the sensitivity Δf is determined, as ε gradually increases, b gradually becomes smaller.

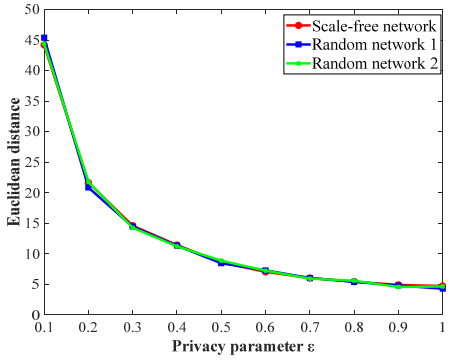

Fig. 3. Euclidean distance under different privacy parameters in the model network.

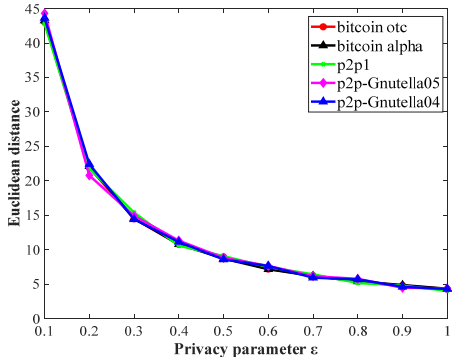

Fig. 4. Euclidean distance under different privacy parameters in the actual network.

The specific Euclidean distance is shown in Table 2 when $\varepsilon = 0.1, 0.3, 0.5, 0.7, 0.9$. When b is large, the overall shape of the probability distribution function of Laplace exhibits a low and wide state, and the opposite is right. The positional parameter is set to 0, as the decrease of b, the probability of occurrence of noise near 0 increases in the generated noise. It means that the difference between the perturbed query result and the original result is getting smaller. Therefore, for the two vectors, the Euclidean distance is getting smaller and smaller, and the practicality of the data is getting larger and larger. As shown in Fig. 4, this conclusion also applies to two actual networks. It can be seen from Figs. 3 and 4 that the Euclidean distance curves in the two model networks are very close. On one hand, the processing of the weight makes the average shortest path length relatively close. Due to the Laplace mechanism in this paper, the two groups of noise generated under the same privacy control parameters are also close. For the above two reasons, the Euclidean distance curves between the two result vectors before and after the perturbation in this experiment are relatively close. The accuracy of the data is retained to 4 digits after the decimal point for accurate analysis.

Table 2. The Euclidean distance of three data sets when $\varepsilon = 0.1, 0.3, 0.5, 0.7, 0.9$.

Data Sets	$\varepsilon = 0.1$	$\varepsilon = 0.3$	$\varepsilon = 0.5$	$\varepsilon = 0.7$	$\varepsilon = 0.9$
p2pGnutella04	43.66	14.41	8.58	5.92	4.61
p2pGnutella05	44.33	14.89	8.81	6.14	4.46
P2p1	42.82	15.40	9.07	6.41	4.76

In summary, the Euclidean distance decreases as the privacy parameter increases, indicating that the data availability in this strategy decreases as the privacy parameter increases. Therefore, in actual demand, if the data still retains large availability after implementing differential privacy, the privacy parameters should not be too large.

(B) Cosine distance

The cosine distance between the original query result vector in the ER and BA network and the result vector after implementing the differential privacy are shown in Fig. 5. The cosine distance between the results of the query before and after the implementation of differential privacy in the two actual networks are shown in Fig. 6.

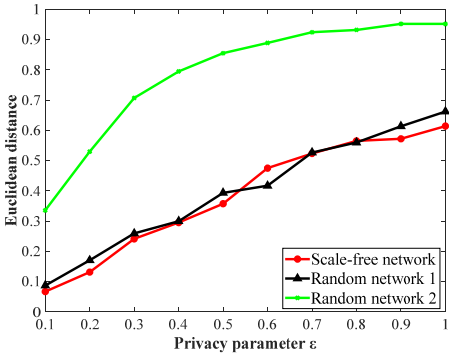

Fig. 5. Cosine distance under different privacy parameters in the model network.

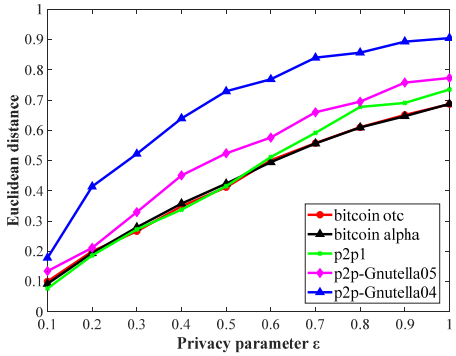

Fig. 6. Cosine distance under different privacy parameters in the actual network.

As can be seen from Figs. 5 and 6, the cosine distance increases as the privacy parameter ε increases. The specific cosine distance is shown in Table 3 when ε= 0.1, 0.3, 0.5, 0.7, 0.9. The probability of noise appearing near 0 increases with the increase of ε. The difference between the vector composed of the query results after differential privacy and the original data is smaller. It indicates that the availability of data at this time is more reserved. Therefore, it is known that after the differential privacy is implemented, the data availability increases as the privacy parameter ε increases. By analyzing the Euclidean distance and the cosine distance under different privacy parameters, the privacy parameter ε should not be controlled too small. The smaller the ε, the harder it is to distinguish the two data sets. When ε is large, although the data availability can be retained, there is no significant meaning for distinguishing the query result from the original data set or the neighbor data set. Therefore, when selecting the parameter ε, the data availability and data security should be considered according to actual needs. If the requirement of data security is high, relatively small ε is selected. If the requirement of data availability is high, relatively large ε is selected.

Table 3. The cos distance of three data sets when ε= 0.1, 0.3, 0.5, 0.7, 0.9.

Data Sets	ε= 0.1	ε= 0.3	ε= 0.5	ε= 0.7	ε= 0.9
p2pGnutella04	0.1783	0.5222	0.7291	0.8403	0.8935
p2pGnutella05	0.1342	0.3293	0.5239	0.6598	0.7581
P2p1	0.0768	0.2725	0.4153	0.5918	0.7350

4. CONCLUSION

The result of the query function in the differential privacy algorithm is sensitive to the scale of the network. This causes the differential privacy execution complexity to vary linearly with network scale changes. In response to this phenomenon, a differential privacy algorithm for querying the average path length (APL) of the network to achieve dual privacy protection of nodes and edge weights in the network is proposed in this paper. According to a typical model network, the global sensitivity of node and edge weight are analyzed and proven. The relationship between data availability and privacy control parameters after the implementation of the algorithm is analyzed experimentally. The algorithm can realize two kinds of privacy protection requirements only through the single-value query of the network APL, and the value itself is not sensitive to the network scale. It can save a large amount of storage space in practical applications.

ACKNOWLEDGEMENT

We would like to thank the anonymous reviewers for their insightful comments.

REFERENCES

1. C. Dwork, "Differential privacy," in *Proceedings of the 33rd International Conference on Automata*, *Languages and Programming*, 2006, pp. 1-12.

2. C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, Vol. 9, 2014, pp. 211-407.
3. X. Xiao, G. Wang, and J. Gehrke, "Differential privacy via wavelet transforms," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, 2011, pp. 1200-1214.
4. R. Sarkar and J. Gao, "Differential forms for target tracking and aggregate queries in distributed networks," *IEEE/ACM Transactions on Networking*, Vol. 21, 2012, pp. 1159-1172.
5. Y. D. Li, Z. J. Zhang, M. Winslett, and Y. Yang, "Compressive mechanism: Utilizing sparse representation in differential privacy," in *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society*, 2011, pp. 177-182.
6. R. Pinot, A. Morvan, F. Yger, C. G. Pailler, and J. Atif, "Graph-based clustering under differential privacy," *arXiv Preprint*, 2018, arXiv:1803.03831.
7. V. Karwa, S. Raskhodnikova, A. Smith, and G. Yaroslavtsev, "Private analysis of graph structure," in *Proceedings of the VLDB Endowment*, Vol. 4, 2011, pp. 1146-1157.
8. V. Karwa, P. N. Krivitsky, and A. B. Slavkovic, "Sharing social network data: differentially private estimation of exponential family random-graph models," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 66, 2017, pp. 481-500.
9. M. Hay, V. Rastogi, G. Miklau, and D. Suciu, "Boosting the accuracy of differentially private histograms through consistency," in *Proceedings of the VLDB Endowment*, Vol. 3, 2010, pp. 1021-1032.
10. R, Chen, B. C. M. Fung, P. S. Yu, and B. C. Desai, "Correlated network data publication via differential privacy," *The VLDB Journal*, Vol. 23, 2014, pp. 653-676.
11. A. Sealfon, "Shortest paths and distances with differential privacy," in *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, 2016, pp. 29-41.
12. S. Brunet, S. Canard, S. Gambs, and B. Olivier, "Edge-calibrated noise for differentially private mechanisms on graphs," in *Proceedings of the 14th Annual Conference on Privacy*, *Security and Trust*, 2016, pp. 42-49.
13. X. Y. Li, J. Yang, Z. L. Sun, and J. P. Zhang, "Differential privacy for edge weights in social networks," *Security and Communication Networks*, Vol. 2017, 2017, pp. 1-10.
14. L. H. Lan and S. G. Ju, "Privacy preserving based on differential privacy for weighted social networks," *Acta Electronica Sinica*, Vol. 36, 2015, pp. 145-159.
15. D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, Vol. 393, 1998, pp. 440-442.
16. J. Wu and J. Watts, "Small worlds: The dynamics of networks between order and randomness," *ACM Sigmod Record*, Vol. 31, 2002, pp. 74-75.
17. Bollobas, *Random Graphs*, Academic Press, London, 1985.
18. Cohen, and S. Havlin, "Scale-free networks are ultrasmall," *Physical Review Letters*, Vol. 90, 2003, pp. 3682-3685.
19. S. Liu, "Global sensitivity analysis: The primer by Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, Stefano Tarantola," *International Statistical Review*, Vol. 76, 2008, pp. 452-452.
20. F. V. D. Heijden, R. P. W. Duin, D. D. Ridder, and D. M. J. Tax, *Classification*, *Parameter Estimation and State Estimation*, Wiley Online Library, 2004.

21. S. H. Lee, P. J. Kim, and H. Jeong, "Statistical properties of sampled networks," *Physical Review E*, Vol. 73, 2006, pp. 016102.
22. B. Ren, D. Guo, Y. Shen, G. Tang, and X. Lin, "Embedding service function tree with minimum cost for NFV-enabled multicast," *IEEE Journal on Selected Areas in Communications*, Vol. 37, 2019, pp. 1085-1097.
23. Xiao, Y. Shen, Y. Zeng, and Y. Zhang, "Cooperative jamming strategy based on community detection for two-hop communication networks," in *Proceedings of IEEE International Conference on Communications*, 2019, pp. 1-6.
24. Y. Zhang, Y. Shen, H. Wang, Y. Zhang, and X. Jiang, "On secure wireless communications for service oriented computing," *IEEE Transactions on Services Computing*, Vol. 11, 2015, pp. 318-328.

Dong Tong (董通) received the bachelor's degree from Beijing Institute of Technology, China, in 2018, and is currently pursuing master's degree at Xidian University, China.

Yong Zeng (曾勇) received his B.Sc, M.S., and Ph.D. degrees from Xidian University in 2000, 2003, and 2008, respectively. Since 2007 he has been with Xidian University as an Associate Professor. His research interests include cryptography, physical-layer security, and complex network.

Zhihong Liu (刘志宏) received his B.Sc. degree from National University of Defense Technology, China, in 1989, his M.S. degree in Computer Science from Air Force Engineering University, China, in 2001, and his Ph.D. degree in Cryptography from Xidian University in 2009. Now he is with the School of Cyber Engineering at Xidian University. His research areas include mobile computing and information security.

Jianfeng Ma (马建峰) received his B.Sc. degree from Shaanxi Normal University, China, in 1985, and his M.Sc. and Ph.D. degrees in Computer Software and Communications Engineering from Xidian University in 1988 and 1995, respectively. He is currently a Professor and Ph.D. Supervisor at the School of Computer Science and Technology, Xidian University. His current research interests include information and network security and computer networks. He has published more than 200 refereed articles in these areas and coauthored more than 10 books. He is a Senior Member of the Chinese Institute of Electronics.

Xiaoyan Zhu (朱晓妍) received the B.E., M.E., and Ph.D. degrees in 2000, 2004, and 2009, respectively, from Xidian University, Xi'an, China. She is currently a Professor with the School of Telecommunications Engineering, Xidian University. Her research interests focus on privacy and security for big data, mobile Internet, cloud computing, online social networks, machine learning, vehicular ad hoc networks, recommendation systems, *etc.*