Journal of Information Security, 2017, 8, 328-338
http://www.scirp.org/journal/jis
ISSN Online: 2153-1242
ISSN Print: 2153-1234

A Dynamic Social Network Data Publishing Algorithm Based on Differential Privacy

Zhenpeng Liu[1,2], Yawei Dong[1], Xuan Zhao[3], Bin Zhang[2*]

[1]School of Cyber Security and Computer, Hebei University, Baoding, China
[2]Center for Information Technology, Hebei University, Baoding, China
[3]School of Electronic Information Engineering, Hebei University, Baoding, China

Email: *zb@hbu.edu.cn

How to cite this paper: Liu, Z.P., Dong, Y.W., Zhao, X. and Zhang, B. (2017) A Dynamic Social Network Data Publishing Algorithm Based on Differential Privacy. *Journal of Information Security*, **8**, 328-338.
https://doi.org/10.4236/jis.2017.84021

Received: September 23, 2017
Accepted: October 17, 2017
Published: October 20, 2017

Abstract

Social network contains the interaction between social members, which constitutes the structure and attribute of social network. The interactive relationship of social network contains a lot of personal privacy information. The direct release of social network data will cause the disclosure of privacy information. Aiming at the dynamic characteristics of social network data release, a new dynamic social network data publishing method based on differential privacy was proposed. This method was consistent with differential privacy. It is named DDPA (Dynamic Differential Privacy Algorithm). DDPA algorithm is an improvement of privacy protection algorithm in static social network data publishing. DDPA adds noise which follows Laplace to network edge weights. DDPA identifies the edge weight information that changes as the number of iterations increases, adding the privacy protection budget. Through experiments on real data sets, the results show that the DDPA algorithm satisfies the user's privacy requirement in social network. DDPA reduces the execution time brought by iterations and reduces the information loss rate of graph structure.

Keywords

Dynamic Social Network, Data Publishing, Differential Privacy

1. Introduction

The innovation of the knowledge society has promoted the advent of the era of "Internet +", such as medical data, big data of intelligent city and large education data, which lead the trend of Internet changes. Social network is a new application mode under the Internet background, and the data dissemination in social network has great research value and application significance. The user's large number of personal privacy information may be leaked when social network data is analyzed and excavated. Social networks are evolving and changing that named dynamic social networks. The dynamic social network is concerned with the dynamic change caused by the change of time in the interaction between social members. The privacy strategy of the static social network data release usually cannot adapt to the dynamic development of social network efficiently. It has far-reaching theoretical significance and practical value in the field of information security and network space security. Existing privacy protection technologies include anonymous technology, data encryption technology, differential privacy technology, privacy information retrieval technology, and accountability system. The social network privacy protection method mainly studies the static social network data dissemination.

2. Related Work

Social network Privacy protection technology mainly includes social network data release privacy protection technology based on clustering, and social network data publishing privacy technology based on the graph modification. Terzi [1] used the degree of the node as the background knowledge of the attacker and proposed the k-degree anonymous attack to prevent the node from being anonymous; Lan Lihui [2] proposed a stochastic perturbation method based on differential privacy model and designed LWSPA algorithm to realize the strong protection of edge and variable weight; Zhang Wei [3] proposed a new algorithm for social network data dissemination which based on the hierarchical random graph and satisfies the difference privacy; Chen Chunling [4] classified the privacy protection level, anonymized the nodes through k-grouping, (k, Δd), and reduced the information loss of the social network graph.

If the static social network data release method is applied directly to the background of the dynamic social network, although it can meet the requirements of privacy protection policy, the time overhead will increase, and the information loss of graph structure will be increased. For the dynamic social network data dissemination method, Ying [5] proposed a random method based on edge spectrum graph to anonymously manipulate the information of nodes and edges, which can prevent attackers from attacking with background knowledge as known condition; Bhagat and Cormode [6] [7] proposed a connection prediction algorithm, which simulates the social network changes caused by new nodes or edges are added to the release graph. The algorithm is based on a predictive chart for group anonymity. Cheng [8] proposed a k-isomorphism privacy protection model. In this method, the iterative distribution of the dynamic social network is processed by the Generalization node identification, which can resist the graph structured attack in the data publishing process;

Karwa and Chen [9] [10] [11] applied the differential privacy technology to the privacy protection of social network. However, the difference privacy complexity of the method graph is high; Guo Caihua [12] summed up the incremental dynamic social network into an incremental sequence of weighted graphs. This paper constructed a privacy protection model based on K-Anonymous weighted graph increment sequence. Guo Caihua proposed an anonymous algorithm WLKA and HVKA based on the weight chain list, which can prevent attackers from attacking based on node label and weight list.

The key of privacy protection problem of dynamic social network data is how to protect the user's sensitive information effectively under the acceptable time cost, and to ensure the loss rate of weight information is small. The main contributions of this paper are as follows.

1) This paper makes an improvement on the social network differential privacy data publishing algorithm based on MCL (Markov Clustering algorithm) [13]. This paper presents the method of data privacy protection which can be applied to the dynamic social network better;

2) In this paper, a strict differential privacy preserving model is introduced. This paper designs a DDPA algorithm that satisfies ε—difference privacy. The DDPA algorithm identifies the edge weight information that changes as the number of iterations increases and adds the privacy protection budget that satisfies ε. The algorithm achieves privacy protection by injecting noise from the Laplace distribution into the weight of the nodes where the nodes are clustered;

3) This paper experiments on the real social network data set. Comparing with the direct application of MDPA algorithm, the DDPA algorithm satisfies the user's privacy requirement in the social network, reduces the execution time and the loss rate of weight information.

3. Basic Concept

3.1. Dynamic Social Network

Definition 1. Dynamic social network

Defining a dynamic social network $G:G=\left(V^I,E^I\right),E^I=\left\{(x,y)\middle|x,y\in V^I\right\}$, V^I represents a collection of users in the social network at the time of iterations, and E^I represents a collection of the edges of the interaction between users in the social network at the time of iterations, $G=\left\{G^1,G^2,G^3,\cdots,G^I\right\}$ represents the collection of social network graphs at $I=1,2,\cdots,N$. $G=\left\{G^{1'},G^{2'},G^{3'},\cdots,G^{I'}\right\}$ represents the collection of social network graphs which has added privacy protection at $I=1,2,\cdots,N$.

The dynamic social Network graph shows in **Figure 1**. $G=(G^{I1}, G^{I2})$. **Figure 1(b)** is compared to **Figure 1(a)**, which increases the edge between node 2, node 6, deletes the edge between node 5 and node 8, and the edge between node 8 and node 11.

3.2. The Edge Weight Information of Ternary Group

Definition 2. The edge weight information of ternary group

Defining a ternary group $T=(i, j, x)$, i, j represents the node number in a social

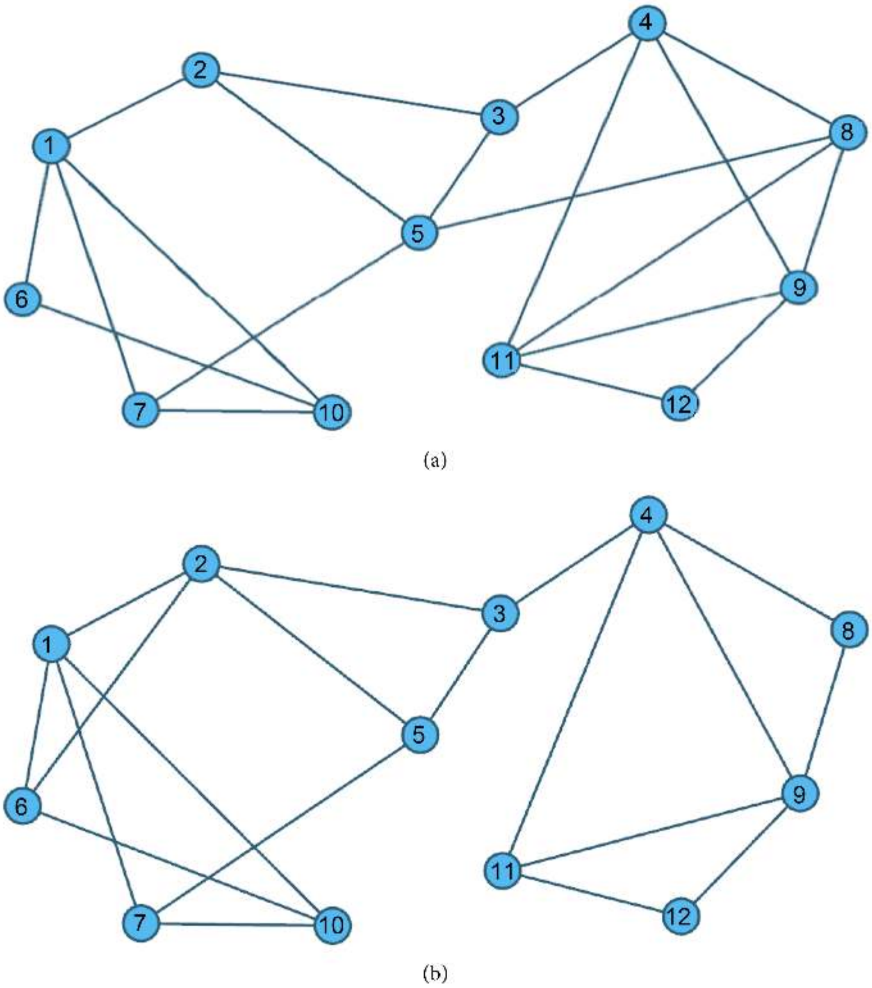


Figure 1. Dynamic social network. (a) Social Network for the first iteration G^{t1}. (b) Social Network for the second iteration G^{t2}.

network diagram, x represents the weight value of the edge. x is 0 when there is no connection between nodes. T = (1, 2, 5) indicates that there is a side between node 1 and node 2, with a weighting value of 5.

3.3. Sensitivity

Definition 3. Sensitivity: Δq is the sensitivity of the query function, which is defined as follows:

$$\Delta q = \max_{D_1,D_2} \left| q(D_1) - q(D_2) \right| \tag{1}$$

Data sets D_1 and D_2 differ by only one element. In this paper, we suppose two social network data sets G^{t1} and G^{t2}. There is only one different element between data sets G^{t1} and G^{t2}. Global sensitivity is set to the maximum difference weight that exists on the difference edge $\Delta f = W_{\max}$.

3.4. *ε*—Weight Vector

Definition 4. *ε*—Weight vector

The social network graph G^n is initialized, and then the Markov clustering is carried out. The clustering result set V of node cluster is obtained, and then the weight information of each cluster is recorded. According to the order of clustering set, the weights are composed of ternary group $T=\{T_1,T_2,\cdots,T_n\}$.

$$\begin{aligned} T_1 &= (i_1, j_1, x_1) \\ T_2 &= (i_2, j_2, x_2) \\ &\vdots \\ T_n &= (i_n, j_n, x_n) \end{aligned} \tag{2}$$

$X=(x_1,x_2,\cdots,x_n)$, according to the query sensitivity Δf and privacy budget parameters ε, constructing a noise vector with a Laplace distribution of length d $<Lap(\Delta f/\varepsilon)>^X$. $P_i = X + Lap(\Delta f/\varepsilon)^X$, P_i is a weight vector satisfying ε—differential privacy.

3.5. Weight Information Loss of Graph

Definition 5. Weight information loss rate of graph

There are $G=(V,E)$ and $G'=(V',E')$. G' is added the privacy protection. The loss rate of weight information due to the change of weight is:

$$WIL(G,G') = \sum_{(i,j)\in E} \frac{|W(i,j) - W'(i,j)|}{W(G)} \tag{3}$$

$W'(i,j)$ is the value of weight which has added the privacy protection. $W(G)$ is the sum of all edge weights of network graphs.

4. A Dynamic Social Network Data Publishing Algorithm Based on Differential Privacy

Applying static social network data privacy publishing algorithm directly to dynamic social network can cause high execution time and large information loss rate of graph structure. This paper makes an improvement on differential privacy network data publishing based on MCL, and designs a dynamic social network data publishing algorithm DDPA which satisfies the ε—difference privacy.

In order to introduce the algorithm flow of DDPA, the MDPA algorithm is decomposed into two parts that include algorithm 1 and algorithm 2.

Algorithm 1: Input the initial social network graph G, expansion parameter e, inflation parameter p, outputs the ε—weight vector of the initial graph G.

Algorithm 2: Output the ε—weight vector of the initial graph G, privacy budget parameters ε, output the privacy preserving graph G'.

4.1. Basic Idea of DDPA

The distribution of social network data has dynamic characteristics, and the graph structure is updated iteratively. DDPA algorithm is an improvement of privacy protection algorithm in static social network data publishing. MDPA algorithm adds noise to the whole network graph, but DDPA algorithm adds noise to the changed network edge weights. DDPA algorithm identifies the edge weight information that changes as the number of iterations increases, and adds the privacy protection budget that satisfies ε. Therefore, DDPA algorithm greatly reduces the execution cost of the algorithm and reduces the loss rate of weight information.

4.2. Basic Flow of DDPA

The algorithm steps are described as follows:

Input: Social Network graph G^I in the Ith Iteration, Social Network graph $G^{I'}$ which has protected in the Ith Iteration, Social Network graph G^{I+1} in the I+1th Iteration, privacy budget parameter ε, expansion parameter e, Inflation parameter p;

Output: Social Network graph $G^{I+1'}$ which has protected in the I+1th Iteration

Step 1 Execute algorithm 1, traverse G^I, build the weight information ternary group $T_I\,(i, j, x)$ and Vector X_I

Step 2 Execute algorithm 2, create a social network graph of privacy protection $G^{I'}$, the weight information ternary group T_I' and Vector X_I' which belong to $G^{I'}$

Step 3 Execute algorithm 1, traverse G^{I+1}, build weights information ternary group $T_{I+1}\,(i, j, x)$ and Vector X_{I+1}

Step 4 Compare T_I and T_{I+1}, recognize ternary group T_c which belongs to modified edges, generate the weight vector X_c corresponding to T_c

Step 5 Compare T_I' and T_{I+1}, recognize ternary group T_a which belongs to add edges, generate the weight vector X_a corresponding to T_a

Step 6 Taking S_i as sampling frequency, make X_c to random sampling. Generating Laplace noise N_c that satisfies differential privacy

Step 7 Taking S_i as sampling frequency, make X_a to random sampling. Generating Laplace noise N_a that satisfies differential privacy

Step 8 Using X_c instead of the changed edge information in the T_I's weight vector X_I', add the edge information increment to X_I', so $X_I' = X_c$

Step 9 According to the query sensitivity Δf and the privacy budget parameter ε, constructing a vector of Laplace distribution with length d: $<Lap(\Delta f/\varepsilon)>^X$

Step 10 Generating a vector $G^{I+1'}$ that satisfies differential privacy:

$$\mathrm{DDPA}\left(G^{I+1'}\right) = P_i = X_I' + Lap\left(\Delta f/\varepsilon\right)^{XI'}$$

Step 11 Distribute social Network graph $G^{I+1'}$, which has protected in the I+1th Iteration

4.3. Privacy Analysis of DDPA

The DDPA algorithm of dynamic social network data release is the improvement of the social network differential privacy data publishing method based on the Markov clustering algorithm in the static social network. The MDPA algorithm has proved that it satisfies ε—difference privacy. This paper only needs to prove that after recognizing the change of the edge weight information, the ε—Weight vector DDPA (G^I) satisfies the differential privacy.

According to the definition of differential privacy, we suppose two dynamic social network data sets G^{I1} and G^{I2}. There is only one different element between data sets G^{I1} and G^{I2}. Given a privacy algorithm DDPA, Range (DDPA) is the range of DDPA. If any outputs of the DDPA algorithm on data sets G^{I1} and G^{I2} satisfy the following inequality, we can say that the DDPA algorithm satisfies ε—differential privacy.

$$\Pr\left[\mathrm{DDPA}\left(G^{I1}\right)\in P_i\right]\le \mathrm{e}^{\varepsilon}\Pr\left[\mathrm{DDPA}\left(G^{I2}\right)\in P_i\right] \tag{4}$$

Proof: Set $p_i \in P_i$, P_i is the same as the X_i dimension. From the conditional probability,

$$\Pr\left[DDPA\left(G^{I1}\right)=p\right]\Big/\Pr\left[DDPA\left(G^{I2}\right)=p\right]$$

$$=\prod_{i=1}^{X}\Pr\left[DDPA\left(G^{I1}\right)_i=p_i\mid p_1,\cdots,p_{i-1}\right]\Big/\Pr\left[DDPA\left(G^{I2}\right)_i=p_i\mid p_1,\cdots,p_{i-1}\right]$$

$$\le\prod_{i=1}^{X}\exp\left\{\left|DDPA\left(G^{I1}\right)_i-DDPA\left(G^{I2}\right)_i\right|\Big/\sigma\right\}$$

$$=\exp\left\{\left\|DDPA\left(G^{I1}\right)_i-DDPA\left(G^{I2}\right)_i\right\|_I\Big/\sigma\right\}$$

$$=\exp\left\{X\left(G^{I1}\right)+Lap\left(\Delta f/\varepsilon\right)-X\left(G^{I2}\right)-Lap\left(\Delta f/\varepsilon\right)\Big/\left(W_{\max}/\varepsilon\right)\right\}$$

$$=\exp\left\{X\left(G^{I1}\right)-X\left(G^{I2}\right)\Big/\left(W_{\max}/\varepsilon\right)\right\}$$

$$\because\left(X\left(G^{I1}\right)-X\left(G^{I2}\right)\right)\le W_{\max}$$

$$\therefore\exp\left\{X\left(G^{I1}\right)-X\left(G^{I2}\right)\Big/\left(W_{\max}/\varepsilon\right)\right\}$$

$$\le\exp\left\{W_{\max}\Big/\left(W_{\max}/\varepsilon\right)\right\}=\exp\{\varepsilon\}=\mathrm{e}^{\varepsilon}$$

$$\Rightarrow\left(\Pr\left[DDPA\left(G^{I1}\right)=p_i\right]\Big/\Pr\left[DDPA\left(G^{I2}\right)=p_i\right]\right)\le\mathrm{e}^{\varepsilon}$$

$$\because p_i\in P_i,\therefore\Pr\left[DDPA\left(G^{I1}\right)\in P_i\right]\Big/\Pr\left[DDPA\left(G^{I2}\right)\in P_i\right]\le\mathrm{e}^{\varepsilon}$$

Then $\Pr\left[DDPA\left(G^{I1}\right)\in P_i\right]\le\mathrm{e}^{\varepsilon}\Pr\left[DDPA\left(G^{I2}\right)\in P_i\right]$

5. Experiment and Analysis

5.1. Experimental Setup

Experimental environment is: Intel(R) Core(TM) i5-4590 CPU @ 3.30 GHz 4.00 GB of Memory. The operating system is Microsoft Windows 7 ultimate. The programming languages are C++ and Matlab. The experimental data is Lesmis which is a weighted social network graph [14]. Lesmis has 77 nodes and 254 sides. In order to reflect the dynamic characteristics of social network, this paper randomly deletes 5% nodes, then adds 5% nodes from the initial social network graph to form a new social network map. In the experimental scheme, the node of social network diagram is invariant, that is, regardless of the node reduction or increase. Only the change of side information is considered.

5.2. Experimental Result

The experiment of this paper contains three parts. The first part of the experiment tests the execution time of the DDPA algorithm. The second part of the experiment tests the graph weight information loss rate of the DDPA algorithm. The third part of the experiment is to compare the DDPA algorithm and the MDPA algorithm in the execution time and the weight information loss rate. The result of the experiment is the average result of five times.

5.2.1. Analysis of Execution Time

The execution time test result sets for the DDPA algorithm are shown in Figure 2.

The experiment tells us the execution time is changing with the change of ε and p. The values of ε are 0.05, 0.1, 1 and 10. At the same iteration time, the increase of ε has less effect on execution time. As the number of iterations increases, the difference edge weight information that needs to be identified during each iteration is reduced. When the ε is invariant, the execution time of the DDPA algorithm is reduced correspondingly.

5.2.2. Analysis of Loss Rate of Weight Information

The test results for the weighted information loss rate of graph in the DDPA algorithm are shown in Figure 3.

The experiment tells us the weighted information loss rate of graph in the DDPA algorithm is changing with the change of ε and p. The values of ε are 0.05, 0.1, 1 and 10. From the experimental results we can see that the weight information loss rate of the graph structure decreases with the increase of ε at the same iteration time. When the value of the privacy budget parameter ε is unchanged, the weight information loss rate of the graph structure increases correspondingly with the increase of the number of iterations. The experimental results show that with the increase of ε, the Laplace noise decreases correspondingly. The value of weights becomes closer to the real value, and then the loss rate of weight information becomes smaller.

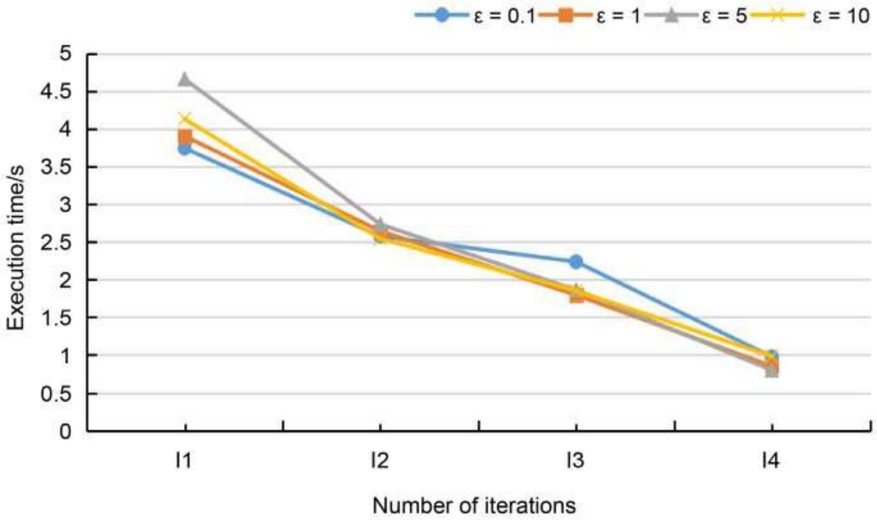


Figure 2. Execution time.

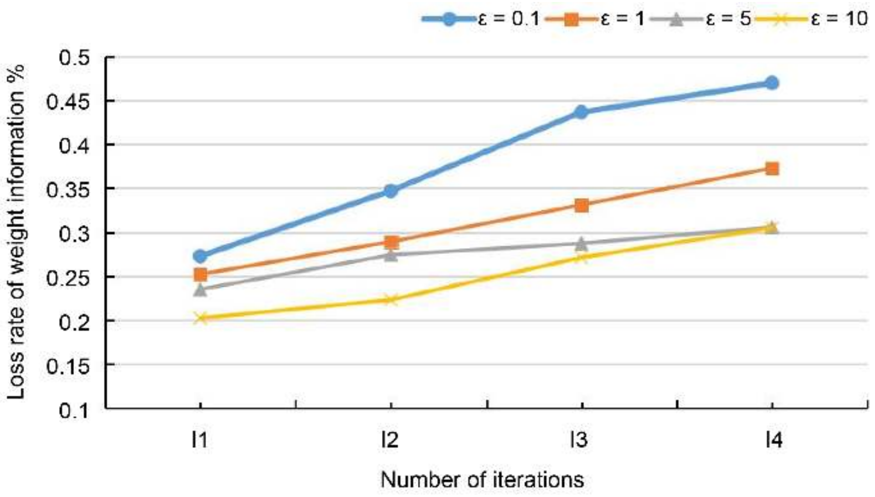


Figure 3. Loss rate of weight information.

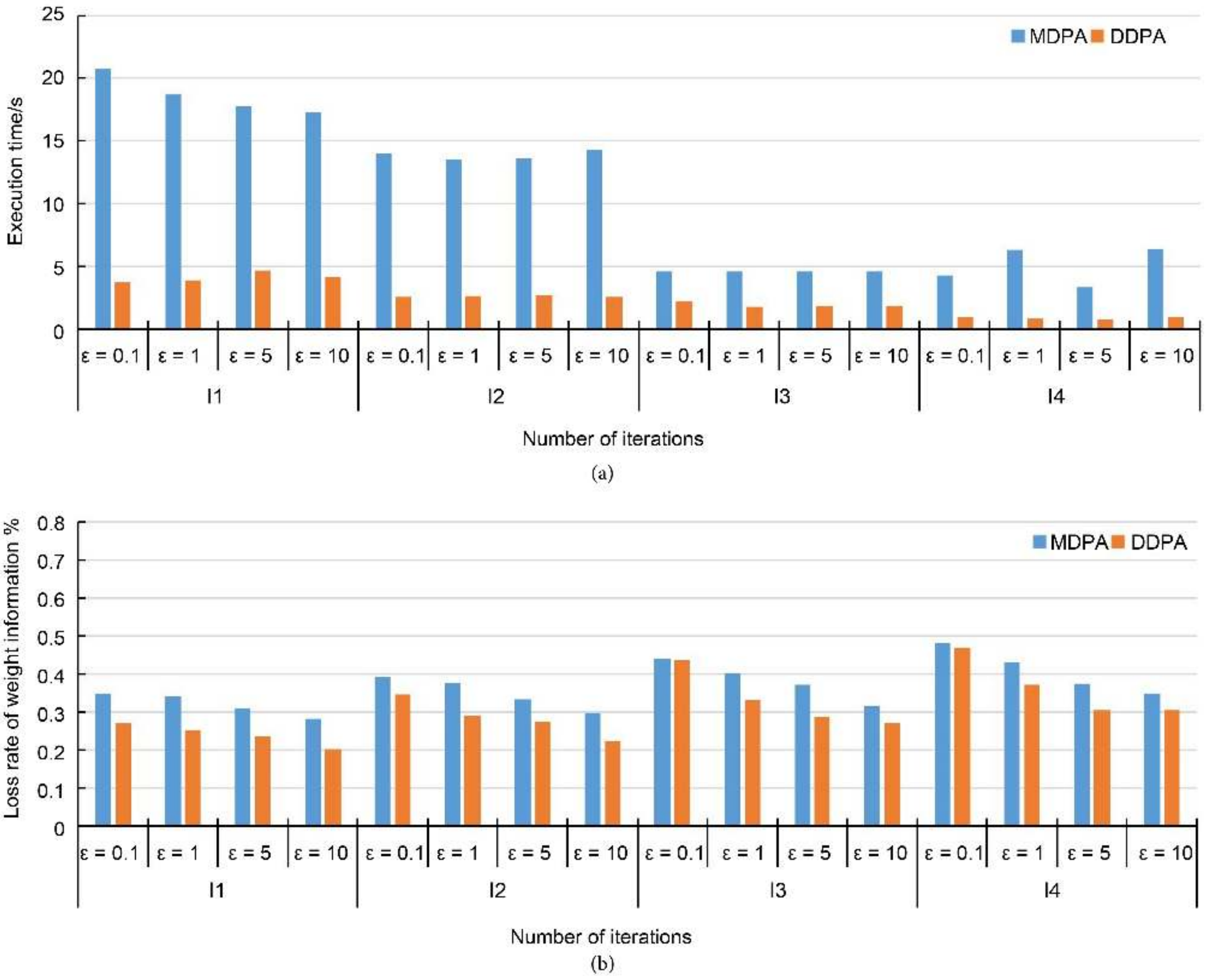
5.2.3. Experimental Comparison

This experiment is a comparison between the MDPA algorithm and the DDPA algorithm in the execution time and the weight information loss rate. The test results are shown in Figure 4.

In the experiment, the values of ε are 0.05, 0.1, 1 and 10. The experimental results in Figure 4(a) show that when the privacy budget parameter ε takes the same value, the execution time of the DDPA algorithm is significantly lower than the MDPA algorithm at the same iteration time, and as the number of iterations increases, the execution time of the DDPA algorithm is also less than the MDPA algorithm. The experimental results in Figure 4(b) show that when the privacy budget parameter ε takes the same value, the weight loss rate of the DDPA algorithm and the MDPA algorithm will increase gradually with the increase of iterative times. Because of the increase of the number of iterations, the Laplace noise satisfying the ε—difference privacy gradually increases, so the weight loss rate will increase correspondingly. However, the weight loss rate of the DDPA algorithm is always lower than the MDPA algorithm, which shows that the DDPA algorithm is superior to the MDPA algorithm in the dynamic social network.

6. Conclusion

Aiming at the improvement of privacy protection algorithm in static social network data publishing, a dynamic social network data publishing algorithm based on differential privacy is designed. This paper recognizes the edge weight information which changes with the increase of the number of iterations, adds the privacy protection budget satisfying ε, reduces the time cost of the algorithm, and guarantees the reduction of the loss rate of weight information. The limitation of this paper is that we only consider the increase or decrease of edge and the change of the edge weight. The change of the node makes the privacy protection budget more complicated. The future work is to deeply study the situation



(a)

(b)

Figure 4. Contrast experiment between DDPA and MDPA. (a) Comparison experiment of DDPA and MDPA in Execution time. (b) Comparison experiment of DDPA and MDPA in Weight information loss rate.

of the change of node. The method of this paper will enhance the degree of privacy protection and reduce the loss rate of weight information under the condition of satisfying the privacy budget.

Acknowledgements

This work has been supported by The Ministry of Education's Research program (2017A20004) and National Science and Technology Support Project (2013BAK07B04).

References

[1] Liu, K. and Terzi, E. (2008) Towards Identity Anonymization on Graphs. *ACM SIGMOD International Conference on Management of Data*, *SIGMOD* 2008, June 2008, Vancouver, BC, Canada, 93-106.

[2] Lan, L.H. and Ju, S.H. (2015) Privacy Preserving Based on Differential Privacy for Weighted Social Networks. *Journal on Communications*, **36**, 145-159.

[3] Zhang, W., Cang, J.Y., Wang, X.R., *et al.* (2016) Differentially Private Network Data Publishing via Hierarchical Random Graph. *Journal of Nanjing University of Posts and Telecommunication* (*Natural Science Edition*), **36**, 23-32.

[4] Chen, C.L., Xiong, J., Chen, L., *et al.* (2016) Personalized Privacy Protection in Dynamic Social Network Data Publication. *Journal of Nanjing University of Posts and Telecommunication* (*Natural Science Edition*), **36**, 74-81.

[5] Ying, X. and Wu, X. (2008) Randomizing Social Networks: a Spectrum Preserving Approach. *Siam International Conference on Data Mining*, *SDM* 2008, 24-26 April 2008, Atlanta, Georgia, USA, DBLP, 739-750.

[6] Bhagat, S., Cormode, G., *et al.* (2010) Privacy in Dynamic Social Networks. *International Conference on World Wide Web*, *WWW* 2010, April 2010, Raleigh, North Carolina, USA, DBLP, 1059-1060.

[7] Bhagat, S., Cormode, G., *et al.* (2010) Prediction Promotes Privacy in Dynamic Social Networks. *Proc of the* 3*rd Conference on Online Social Networks*, *CA*: *USENIX Association*, **2010**, 6.

[8] Cheng, J., Fu, W.C., Liu, J. (2010) K-Isomorphism, Privacy Preserving Network Publication against Structural Attacks. *ACM SIGMOD International Conference on Management of Data*, *SIGMOD* 2010, June 2010, Indianapolis, Indiana, USA, DBLP, 459-470.

[9] Karwa, V., Smith, A., *et al.* (2014) Private Analysis of Graph Structure. *ACM Transactions on Database Systems*, **39**, 1146-1157. https://doi.org/10.1145/2611523

[10] Chen, S. and Zhou, S. (2013) Recursive Mechanism: Towards Node Differential Privacy and Unrestricted Joins. *Proc of the International Conference on Management of Data*, **2013**, 653-664.

[11] Chen, R., Fung, B.C.M., Yu, P.S., *et al.* (2014) Correlated Network Data Publication via Differential Privacy. *The VLDB Journal*, **23**, 653-676. https://doi.org/10.1007/s00778-013-0344-8

[12] Guo, C.H., Wang, B., Zhu, H.J., *et al.* (2016) Incremental Dynamic Social Network Anonymity Technology. Journal of Computer Research and Development, **53**, 1352-1364.

[13] Liu, Z.P., Dong, Y.W., *et al.* MDPA: Differential Privacy Network Data Publishing Based on MCL. *Journal of Zhengzhou University*, In Press.

[14] Knuth, D.E. (1993) The Stanford GraphBase: A Platform for Combinatorial Computing. Addison-Wesley, Reading, MA.