Multi-mode Network Analysis under Differential Privacy

Yuning Song¹, Liping Ding¹,²*, Mengying Dong³, Xuehua Liu¹, Xiao Wang³

¹(Laboratory of Parallel Software and Computational Science, Institute of Software, Chinese Academy of Sciences, Beijing 100190)
²(Digital Forensics Laboratory, Institute of Software Application Technology, Guangzhou (GZIS), Guangzhou 511458)
³(State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093)

Corresponding author’s e-mail: dingliping@gz.iscas.ac.cn

Abstract With the advent of the big data era and the advancement of social network analysis, the public is increasingly concerned about the privacy protection in today’s complex social networks. For the past few years, the rapid development of differential privacy (DP) technology, as a method with a reliable theoretical basis, can effectively solve the key problem of how to "disassociate" personal information in social networks. This paper focuses on the multi-mode heterogeneous network model which has attracted a lot of attention in the field of network research. It introduces differential privacy and its application in big social networks briefly first, and then proposes a centrality-analysis method based on DP in a typical social network, i.e. the multi-mode network. The calculation principle and applicable scenarios are discussed. Then, its utility is analyzed and evaluated through experimental simulation. Possible improvement of DP algorithm in multi-mode networks above is prospected in the end.

1. Introduction
Driven by mobile Internet technology, people have gradually got used to using Internet for everything. Advances in data mining and machine learning technologies have enabled the information of social activities to be gathered into various big data systems. That is, the government, research institutes and commercial organizations are striving to obtain user information to help enhance service quality or get business benefits.

In social networks, individuals are connected as groups, and groups converge into bigger ones, which are even more complicate when groups of different sizes and various connection relationships are involved. The development of modern big data-based network analysis techniques has made modeling and analysis of such complex social networks possible. From simple commercial recommendation systems to sophisticated social network-based user profiling, knowledge discovery acts on almost all traces of people’s online activities in modern society[1]; and therefore, the privacy protection has become a serious problem.

Nowadays, when traditional anonymization techniques are becoming gradually less capable to cope with advanced big data mining, the emergence and development of differential privacy (DP) technology provide a timely and powerful means for privacy security in social networks of big data. Its key role is to hide the information about the existence of the individual in terms of statistical significance from the global analysis results of big data[2]. The strict definition of privacy protection
level has led DP into the practical application in a growing number of fields, especially in the study of big data social networks, which are more sensitive to privacy issues. It must be acknowledged, however, that the more complex the structure of the network is, the more difficult it is to achieve strict privacy protection. For an individual in a big network carrying a large amount of relational information, generally it is very difficult to hide its “existence” and ensuring the accuracy of data analysis results at the same time[3]. Current approaches to the privacy protection of a network structure, especially for nodes of it, usually rely on adding preconditions to networks parameters or query patterns to improve the utility of the results[4-7], in which factors more influential to the analysis output are often limited by strong constrains, with the cost of narrowing the “width” of statistical query.

The main goal of this paper is to relax some restrictions on data analysis targets to more complex statistical quantities in “multi-mode networks” with the property of multi-layer, rather than expecting a universal method that can be applied to all kinds of networks. We explore how to effectively solve the privacy protection problem of individual nodes in multi-mode networks with DP. By analyzing the key features of such networks, some statistical problems can be reasonably transformed and refined; by optimization of DP perturbation deduced from hierarchical relationships in multi-mode networks, we propose a class of privacy-preserving methods that meet the requirement of strict differential privacy definition.

The paper is organized as follows: the second section provides an overview of the basic principles of differential privacy and its application in network analysis, as well as the current research progress in this field; the third section first introduces multi-mode network, including its key issues and main difficulties for DP, then proposes a privacy-preserving algorithm for the statistics of centrality based on DP, using a two-mode network as an example, and finally discusses how to extend it to multi-mode network scenarios; the fourth section analyzes the performance of this DP method as well as its constrains through experimental simulation; the last section discusses some promising research directions for the application of DP methods in complex network privacy problems based on the summary.

2. Differential Privacy and Social Network Analysis

2.1 Differential Privacy Fundamental

Differential privacy (DP) is a privacy-preserving method based on data distortion, which is accomplished by differential perturbation on the output of queries to the data set[8]. If a query function $F$ on two data sets $D_1$ and $D_2$ with arbitrary differences of at most one data item can satisfy

$$\frac{P[F(D_1) \in S]}{P[F(D_2) \in S]} \leq \exp(\varepsilon) \tag{1}$$

We say $F$ satisfies $\varepsilon$-DP, while in Equation (1), $P[F(D_i) \in S]$ is the probability of $F(D_i) \in S$, and $\varepsilon$ is the privacy budget, which is used to measure the degree of privacy protection.

For network structure analysis, we use the most conventional way to describe the network as a graph. For graph $G$, $(V, E)$ represent the set of all nodes and edges in the graph, which means $(V, E)$ carries all the information of the graph. Assuming $G$ is an undirected and unweighted graph, according to the basic definition of DP, by selecting $E$ and $V$ as object data sets respectively, we give two kinds of definitions of DP in $G$.

1) For any $G$, $(V, E)$ and $G'$, $(V', E')$, if the difference between $(V, E)$ and $(V', E')$ is only one edge. that is, $E$ can be converted to $E'$ by adding or removing one edge, then if

$$\frac{P[F(G) \in S]}{P[F(G') \in S]} \leq \exp(\varepsilon) \tag{2}$$
is satisfied, We say $F$ satisfies $\varepsilon$-edge DP[9].

2) For any $G$, $(V,E)$ and $G'$, $(V',E')$, if the difference between $(V,E)$ and $(V',E')$ is only one node (vertex), that is, $E$ can be converted to $E'$ by adding or removing one vertex and all edges connecting to it, then if

$$\frac{P[F(G) \in S]}{P[F(G') \in S]} \leq \exp(\varepsilon)$$

(3)

is satisfied, We say $F$ satisfies $\varepsilon$-node DP[4].

Function $F$ acting on the graph $G$ is a statistical function with randomized output, which is usually constructed by adding differential noise on the original one[10].

$$F(G) = f(G) + N(\Delta f / \varepsilon)$$

(4)

$f$ is the original function. $N(\Delta f / \varepsilon)$ represents a random noise function and the amplitude of the randomness is determined by $\Delta f / \varepsilon$. Here $\Delta f$ is the sensitivity of $f$, representing its requirement for randomness.

$$\Delta f = \max_{G,G'} \| f(G) - f(G') \|$$

(5)

if no restriction for $G$ and $G'$ in equation (5), we call $\Delta f$ the global sensitivity of $f$ [11]. From the above definitions, we can find that for general $f$, the sensitivity of node DP is significantly greater than that of edge DP.

2.2 Research on differential privacy in social network analysis

In classical network studies, researchers focus on regression analysis of individual information to get useful patterns, while in the case of social networks, the network structure formed by connections between individuals becomes the focus. In general social network models, nodes often correspond directly to natural individuals, and edges represent connections between them. There are lots of kinds of a connection relationship, like a calling in telecom network, an Email forwarding on the internet, or a “follow” in social software.

The purpose of DP in social network models is to ensure that the statistical results obtained from social networks do not reveal any information about individuals or inter-individual relationships by the introduction of differential noise perturbation.

On DP for protecting relationship privacy, as Equation (2) above, researchers have had extensive finds. From queries to node degree of social networks[13], solving and publishing of network adjacency matrix[14], shortest path in weighted networks[9], cut sets[6], to parameter fitting of network models[15,16], structure queries[17], and even more like data mining for social networks, including frequent subgraph[18], classification[19], clustering[20] and so on, there has been a great deal of theoretical analysis and empirical exploration.

When it comes to DP for protecting individual privacy, as Equation (3) above, for most of network analysis methods, the high noise amount caused by the high sensitivity of nodes will greatly affect analysis results, thus the research on DP for nodes is still in progress. Nissim improves the accuracy of the results of node DP by projection methods[4]. Day and Macwan propose a degree publishing method with the help of similar ideas[21,22]. For the basic Erdos-Renyi random graph model, Ullman proposed a parameter estimation method that satisfies node DP[23]. Chen optimize noise in network structure queries using “empirical sensitivity”[24]. In comparison to edge DP, node DP often require very strong constrains and some approximations in exchange for an acceptable results.

2.3 Differential privacy and multi-mode networks

A multi-mode network is a network with multiple classes of nodes (and multiple classes of edges), and the relationship between different classes of nodes are often more important than those of the same class. The most common multi-model network model in social networks is the affiliation network (two-mode network), which represents a kind of subordination relation. In fact, corporation staff networks, user recommendation networks, online community topic networks, etc. can be abstracted as
two-mode or multi-mode networks. The following diagram shows a simple course selection network (two-mode network).

![Figure 1. Teachers-Students Network](image)

We propose a DP method to achieve some available statistical queries with the help of features of multi-mode network in this paper. To balance the complexity of differential noise and to reflect key characteristics of the network as more as possible, we need to choose suitable statistical objects to ensure the results have both accuracy and practical research value. The most important purpose of social network research is that information beyond independent attributes of nodes can be uncovered through the relational information. One of the key properties of multi-mode network structures is how the influence of a node is usually implied in the connectivity of the nodes, which is generally summarized as “centrality” in most studies. In fact, centrality is one of the most fundamental and important properties of nodes in social networks[3], so we will discuss some centralities as following later.

1) degree centrality: the degree centrality $D_i$ of node $i$ is determined by the number of connections of the node, that is, $D_i = \text{deg}_i$.

2) closeness centrality[25]: the closeness centrality $C_i$ of a node is usually used to describe the average distance to all other nodes. $C_i = \frac{1}{(n-1)} \sum_{i \neq j} d_{ij}$, $d_{ij}$ is the shortest path between node $i$ and $j$.

3) betweenness centrality[26]: the betweenness centrality $B_i$ of node $i$ is used to measure the role it plays in the whole connectivity of the network. $B_i = \sum_{j,k} g_{jk} / l_{jk}$, $l_{jk}$ is the number of shortest paths between $j$ and $k$, while $g_{jk}$ is the number of route nodes.

Centrality is chosen as the target of analysis in this paper because it is often most directly exposed and concerned in complex network structure. For example, a hot character embodies degree centrality; group organizers usually have the highest closeness centrality; a commonly used automated email copy list may be a reflection of betweenness centrality. The purpose of centrality computing is to help researchers determine the influence of nodes, such as which nodes are most critical to the connectivity of the whole network or the dissemination of messages.

### 3. Differential Privacy in Multi-mode Networks

#### 3.1 Differential Privacy of Centrality

As mentioned above, in social network models, nodes correspond to individuals. For the whole network, the output which complies with DP should be independent of the any individuals in a probabilistic sense. However, it requires a great perturbation of the output to achieve this. Taking degree centrality in 2.2 as an example, the output of such information can be generalized to the following form,

$$f(G) = f(D_1, D_2, ..., D_n) \tag{6}$$

For degree centrality, expanding (6) and ignoring higher order terms, we approximately have
Due to the arbitrary selection of nodes, the amplitude of DP noise to be added is actually proportional to \( \max\{k_i\} \propto n \), which means the noise required for node DP without any constraint is large.

For closeness centrality, \( C_i \) can also be processed to a form similar to (7). For betweenness centrality, it can be explained as follows. An arbitrary edge node in a relatively sparse network, whose betweenness is 0, can obtain an optimum close to 1, by adding a new node bridging all other nodes simply, and the path of length 3 between the original edge node and the newly added node becomes the shortest path between most nodes. The Network \( G \) becomes \( G' \) after adding new nodes as above, and for any node \( i \) we have

\[
f(G) - f(G') \approx k_i (B - B') \approx k_i
\]

(8)

The noise required is still determined by \( \max\{k_i\} \propto n \). Although strict DP can be achieved in this case, the noise in the results is of the same magnitude as the original results, and the accuracy of the output suffers a large loss.

One of the optional solutions to the above problem is to sacrifice some privacy to ensure the accuracy of the results, which is called privacy degradation[24]. If equation (1) is relaxed to

\[
P[F(D_a) \in S] \leq \exp(\epsilon) * P[F(D_B) \in S] + \delta
\]

(9)

then \( F \) satisfies \( \epsilon - \delta \) differential privacy. From a dataset perspective, \( \delta \) represents relaxing the privacy at some nodes with greater local sensitivity, i.e., allowing privacy degradation at nodes with greater impact on the output[27].

If we simply concern about the edge DP, which means setting protection objects as relationships, for (7) and (8), the difference between \( G \) and \( G' \) is only one edge. Thus the noise still depends on \( k_i \), but \( k_i \) does not increase as \( n \). That is, for centrality or more generalized statistic, the noise required to maintain edge DP does not increase with network size.

### 3.2 Multi-mode Networks

As in 3.1, \( \epsilon - \delta \) DP is in fact a concession to the privacy of some nodes. Therefore if the target of node DP is limited to some nodes, the privacy degradation problem is evaded. We make this prerequisite satisfied with the help of multi-mode networks. For social network models, multi-mode means there are nodes with completely different properties in the network[28], and we focus on such networks.

First, without loss of generality, we consider two-mode networks, which contains two kinds of nodes \( \{A, B\} \), node \( a \in A \), and node \( b \in B \). According to the analysis in 3.1, assuming the connection between \( B \) and \( A \) represents a partial order relation “\( \prec \)”, we consider privacy protection of \( A \).

Here we define the centrality of one kind of node based on the partial order relation through the whole network model.

**Definition 1.** if for any \( a \in A, b_1, b_2 \in B \), \( b_1 \prec a \) and \( b_2 \prec a \) are both satisfied, then there is an edge between \( b_1 \) and \( b_2 \) in \( \{B\} \), and the weight is proportional to the number of \( a \) satisfying the conditions.

For centrality in 2.2, if only edges of definition 1 are considered, we obtain the centrality of a single kind of nodes. In social networks modeled as the multi-mode network, such definition is just of interest to most researchers. For example, if we use \( A \) for search service users, \( B \) for a collection of hot search, \( \prec \) for search behavior, then the “relationship / edge” represents a common search tendency, which is more useful than search behavior itself. Or \( A \) for customers, \( B \) for products, \( \prec \) for buying behavior, we obtain an edge for a common shopping tendency. It corresponds exactly to what centrality actually represents, i.e., the strength of association between nodes of the same kind. In
this two-mode scenario, we may refer to \{A\} as the first-level network (which is, the main group facing privacy problems), \{B\} as the second-level network.

Obviously at the beginning, we only have the original network consisting of edges represented by the partial order relations, and it can be converted to the network in Definition 1 in the following way. Adjacency matrix \(M\) represents the network formed by the relationship between \(A\) and \(B\).

\[
M = \begin{pmatrix}
m_{11} & \ldots & m_{1c} \\
\vdots & \ddots & \vdots \\
m_{n1} & \ldots & m_{nc}
\end{pmatrix}
\]  

(10)

\(m_{ij} = 1\) means the edge \(b_i \prec a_j\) exists, then we could obtain the network \{B\} formed by nodes of \(B\) whose adjacency matrix is \(M_B\),

\[
M_B = MM^T
\]  

(11)

Reconsidering the effect of one node existence of \(A\) on centrality of \(M_B\) reveals that its contribution is limited to unit weight of edges in \(M_B\).

It is important to note that the above analysis separates the objects of privacy protection (natural persons in most cases) and the object of statistical queries, which make it easier to achieve node DP, and can also output valuable statistical results at the same time. Here the centrality of second-level network \{B\} is actually clustering centrality, which means the more similar one is to others, the more central it is. The most common and typical example is the network of recommendation system.

3.3 Differential Privacy for two-mode Networks

According to methods in 3.2, by the construction of second-level network and clarify the definition of its centrality, the amount of noise in node DP can be compressed. Next we consider each of the three centralities in 2.2.

For degree centrality, the sensitivity of DP algorithm is intuitive. In fact, the local sensitivity of the degree distribution function is 1 for any node of unweighted original networks, since a single node can contribute at most one connection to the corresponding network.

As for closeness and betweenness, although explicitly defined, in practice they are usually not computed using an ergodic approach. For example, to calculate betweenness, the following approach of approximation is often used as a prototype algorithm[29, 30]:

Table 1. Calculation of betweenness centrality.

| Calculating the betweenness centrality \(B\) of node \(a\) |
|----------------------------------------------------------|
| 1: random choose subset \(g\) from network \(G\)          |
| 2: for \(B(a) = 0, a \in g\)                             |
| a) init empty stack \(S\), empty pre-order node list \(p(a)\), empty breadth-first \(q\); |
| b) \(a\) enqueues into \(g\)                             |
| c) while \(g\) not empty                                 |
| i. \(b\) dequeues from \(g\), push to \(S\)             |
| ii. for \(c\) is neighbor of \(b\)                      |
| if distance between \(a\) and \(c\) is first computed, then \(c\) enqueues \(q\), compute shortest path \(r(c)\) between \(a\) and \(c\), if \(b\) is in \(r(c)\), then \(c\) enqueues into \(p(a)\), compute the number of shortest paths \(w(c)\) |
| d) init all dependent variables \(r\) of \(a\) to 0       |
| e) while \(S\) not empty                                 |
| i. \(S\) pops \(d\)                                    |
| ii. for pre-order list of \(d\) compute corresponding \(r\) |
| iii. accumulate \(r\) to \(B(a)\)                       |
| 3: output \(B(a)\), \(a \in g\)                        |
We note that the first step of the above algorithm is to pick a portion of nodes from the entire network to reconstitute a subnet, and approximate the true result in the original network by calculating the betweenness centrality in the subnet. Generally nodes can be selected using specific probabilities or according to the degree of the node. If selected with a specific equal probability to form a subnet, which might cause a longer iteration, it is actually equivalent to using the random response method of DP[31]. Assuming the specific probability is \( p \), from (4) and (5) it can be derived that the process of constructing subnet make the centrality results satisfy the DP with privacy budget as

\[
\varepsilon = \log(\max(\frac{p}{1-p}, \frac{1-p}{p}))
\]

(12)

Furthermore, the method of selecting nodes randomly to form a subnet can actually be directly applied to the calculation of degree and closeness centrality, thus resulting in a unified approach to the calculation of different centrality. The goals of reducing the whole computational complexity and introducing DP mechanism are accomplished simultaneously in the process of constructing the subnet.

If computational complexity of betweenness is considered first, the process of subnet construction usually require a degree threshold for rapid convergence. We can ignore that issue since from the perspective of DP, different randomness between nodes will introduce privacy degradation for an overall privacy budget.

3.4 Extension to Multi-mode networks
Looking back at Definition 1, we note that in constructing the network \( \{B\} \), a pair of partial order relations in original \( \{A,B\} \) actually constitute one edge of unit weight in \( \{B\} \), which means in \( \{A,B\} \rightarrow \{B\} \), the whole connectivity is exponentially decaying. The transformation of node DP in 3.2 is based on this change, and finally reduces the whole sensitivity. Similarly, we can define hierarchical network models of multi-mode networks. Taking the three-mode network as an example:

**Definition 2.** if for any \( a \in A \), \( b_1, b_2 \in B \), \( c_1, c_2 \in C \), \( b_1 < a \), \( b_2 < b_1 \) and \( c_1 < b_1 \), \( c_1 < b_2 \), \( c_2 < b_1 \), \( c_2 < b_2 \) are all satisfied, \( \prec \) and \( \prec \) represent partial order relations between \( \{A,B\} \) and \( \{B,C\} \) respectively, then there is a edge between \( c_1 \) and \( c_2 \), and the weight is proportional to the number of \( \{a,b_1,b_2\} \) tuples satisfying the conditions.

There are many examples of social networks for the above definition of networks. For example, if we use \( A \) for the audience, \( B \) for a collection film review site, \( C \) for movies, \( \prec \) and \( \prec \) for review actions.

From Definition 2, one edge in \( \{C\} \) requires 4 edges in \( \{B,C\} \). The decay of connections makes it easier to preserve the privacy of \( \{A\} \) in \( \{C\} \) than in \( \{B\} \), which will be verified in the experiments in the next section.

4. Experimental evaluation

4.1 Experimental environment and datasets
The experimental in this section will implement the algorithm in Section 3 with open source python (ver. 3.6) package NetworkX (ver. 2.5). The experimental environment is a SuperVessel host, loaded with the image of Python Science Computing and Application Development v2.2.0 on ubuntu 16.04LST 64bit, vCpu 4core@3.0 GHz/16G Ram.

Limited by the computational complexity and memory limitation of betweenness centrality itself, a sparse two-mode network dataset of medium size has been chosen for experiments[32]. The 2054 nodes of the first-level network are natural individuals (voters); the 29 nodes of the second-level network are groups (candidates); the number of edges is 18101, and the network density is 0.0083.
4.2 Experimental methods and Evaluation
Due to the constraints of the experimental environment and the efficiency of IPython, the absolute speed of the algorithm is not concerned. In fact, the size of the object dataset (relating to whether swap partition is used) and the efficiency of NetworkX API are the main factors limiting the speed of the simulation.

As the principle of the algorithm has been analyzed in 3.2 and 3.3, this section focus on the utility of different centrality calculation results after applying random response mechanism, that is, how much the DP mechanism will affect the accuracy of centrality analysis results.

Considering that the centrality is first a property of individual nodes, while what is required to be evaluate in this section is the impact of DP on the centrality analysis of the whole network, it is important to have a single metric capable of evaluating the overall method. Since centrality is a metric used to characterize the importance of nodes for the whole network, the ranking of node centrality is actually a single scalar indicator that is often used in practical social network research. The relationship between the ranking results (Rank correlation) obtained by introducing DP can be used to evaluate the utility of the algorithm results[33]. The number of normalized reverse order pairs (Kendall Tau factor, KT factor) is used as the quantification of the utility[34].

\[ \tau = \frac{2(R - W)}{n(n-1)} \]  

\( n \) is the number of nodes, \( R \) and \( W \) are the numbers of positive and reverse order pairs in the corresponding sorting results, respectively.

4.3 Experimental results and analysis
According to the methods described in 3.1 and 3.2, the utility of centrality statistics is examined under different DP budget usage, where calculation results of the three kind of centrality are the average of 100 experiments to reduce random errors. As a comparison, the mainstream projection-based algorithm is chosen for degree and closeness centrality[21,22]. However, for betweenness, there is no other dedicated algorithm based on DP, so K-path which uses noise deduced by global sensitivity with high calculation efficiency[30] is used for an approximate comparison.

![Figure 2. DP centrality results of 2-mode network](image-url)
As shown in Figure 2, all three centrality results have good utility on a high DP budget. For degree and closeness, our algorithm is better than the comparison one; the advantage of betweenness is more significant because the comparison algorithm uses global sensitivity. When the privacy budget is low, the relative errors increase rapidly, especially for closeness and betweenness. The reason is that the experimental dataset is real two-mode network data with properties of strong scale-free[35], which means the distribution of the connectivity nodes has the long-tail effect that is very common in reality. Centralized connections make most nodes a low centrality, leading to less difference between comparison algorithm and ours.

In studies of practical social network, nodes with low centrality tend to get low attention, so ones with very low centrality can be ignored. If the utility calculation is restricted to nodes with high centrality, which means great influence on the network structure, and the result is shown in Figure 3. All nodes are counted in while the computation of equation (13) only considered the top 30% nodes of higher centrality.

![Figure 3. DP centrality results of 2-mode network (top 30% nodes)](image)

It can be seen that if we only focus on nodes with higher centrality, the statistics under the same DP method can have significantly better accuracy.

As discussed in 3.4, our DP method is applied to a three-mode network, which is constructed by re-slicing the first-level network according to the family affiliation relationship between natural persons. The numbers of nodes in two new level are 1602 and 452, resulting in a very big adjacency matrix in the calculation. This three-mode network is just constructed for comparison with the two-mode one, and the purpose is to study the interaction between node hierarchy and DP randomness. The experiment also restricts the nodes of interest to top 30%.
As shown in Fig.4, the overall utility of the results has improved in three-mode networks, which is consistent with the theoretical analysis in 3.4. In fact, the definition of multi-mode networks in practical social network research is much broader, and nodes can be different classes of elements in social activities. For example, a natural three-mode network describing the organizational structure can be formed by employees, project groups, and departmental leads.

5. Conclusion
In this paper, from the perspective of practical studies we propose a method that can apply differential privacy protection to node privacy in multi-mode networks. For the structural characteristics of such complex networks, we analyze the principle and effect of the method theoretically, design the whole algorithm of centrality based on DP, and then give the experimental method and result evaluation.

There have been numerous methods and tricks for researchers to apply DP techniques to the study of social networks. Due to the impact of DP on network structure, the current approach focuses on mainly the protection of relation information; for privacy of individual nodes in a network, DP algorithm is often implemented to get a better utility by adding restrictions. The method of this paper is actually based on this idea: by analyzing the essential characteristic of multi-mode networks, we choose to limit the protection to one class of nodes among them whose privacy is vulnerable. By combining the randomness of centrality calculation and DP random response mechanism, we have unified DP algorithm for different statistics and further improved the accuracy of statistical results. However, there is much work to be done in the future on how to loosen the current constraint on protection targets and scenarios to enable node DP methods to be applied to a broader range of networks.

Acknowledgments
This work has been supported by the 2019 Artificial Intelligence Application Demonstration Project of Nansha District (No.2019SF01).
References

[1] Michael, G., Le, G. (1999) A survey of data mining and knowledge discovery software tools. ACM SIGKDD explorations newsletter, 1(1): 20-33.

[2] Task, C., Clifton, C. (2012) A Guide to Differential Privacy Theory in Social Network Analysis. In: Advances in Social Networks Analysis and Mining. Istanbul. pp. 411-417.

[3] Goodreau, S. M., Kitts, J. A., Morris, M. (2009) Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. Demography, 46(1): 103-125.

[4] Kasiviswanathan, S. P., Nissim, K., Raskhodnikova, S., et al. (2013) Analyzing Graphs with Node Differential Privacy. In: Theory of Cryptography Conference. Tokyo. pp. 457-476.

[5] Hay, M., Li, C., Miklau, G., et al. (2009) Accurate Estimation of the Degree Distribution of Private Networks. In: International Conference on Data Mining. Miami. pp. 169-178.

[6] Upadhyay, J. (2013) Random Projections, Graph Sparsification, and Differential Privacy. In: International Cryptology Conference. Santa Barbara. pp. 276-295.

[7] Sala, A., Zhao, X., Wilson, C., et al. (2011) Sharing Graphs Using Differentially Private Graph Models. In: Acm Sigcomm Conference on Internet Measurement Conference. Chicago. pp. 81-98.

[8] Dwork, C. (2006) Differential Privacy. In: International Colloquium on Automata Languages and Programming. New York. pp. 1-12.

[9] Costea, S., Barbu, M., Rughinis, R. (2013) Qualitative Analysis of Differential Privacy Applied over Graph Structures. In: 11th RoEduNet International Conference. Romania. pp. 1-4.

[10] Dwork, C. (2011) A firm foundation for private data analysis. Communications of the ACM, 54(1): 86.

[11] Dwork, C., Mcsherry, F., Nissim, K., et al. (2006) Calibrating Noise to Sensitivity in Private Data Analysis[C]. In: Th eory of Cryptography Conference. New York. pp. 265-284.

[12] Goldenberg, A., Zheng, A.X., Fienberg, S. E., et al. (2009) A Survey of Statistical Network Models. Foundations and Trends in Machine Learning, 2(2): 129-233.

[13] Karwa, V., Slavkovic, A.B. (2012) Differentially Private Graphical Degree Sequences and Synthetic Graphs. In: Privacy in Statistical Databases. Corfu. pp. 273-285.

[14] Ahmed, F., Jin, R., Liu, A. X. (2013) A Random Matrix Approach to Differential Privacy and Structure Preserved Social Network Graph Publishing. arXiv: Cryptography and Security.

[15] Wang, Y., Wu, X., Wu, L. (2013) Differential Privacy Preserving Spectral Graph Analysis. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. Gold Coast. pp. 329-340.

[16] Karwa, V., Krivitsky, P.N., Slavkovic, A.B. (2017) Sharing social network data: differentially private estimation of exponential-family random graph models. Journal of the Royal Statistical Society, 66(Pt.3), 481-500.

[17] Karwa, V., Raskhodnikova, S., Smith, A. D., et al. (2011) Private analysis of graph structure. PVLDB, 4(11): 1146-1157.

[18] Shen, E., Yu, T. (2013) Mining Frequent Graph Patterns with Differential Privacy. In: ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Chicago. pp. 545-553.

[19] Johansson, F. D., Frost, O., Retzner, C. T., et al. (2015) Classifying Large Graphs with Differential Privacy[C]. In: Modeling Decisions for Artificial Intelligence. Skövde. pp. 3-17.
[20] Wang, X., Li, Y., Jin, Y., et al. (2019) Spectrum-Centric Differential Privacy for Hypergraph Spectral Clustering. Parallel and Distributed Computing, Applications and Technologies. Springer, Singapore.

[21] Day, W.Y., Li, N., Lyu, M.. (2016) Publishing Graph Degree Distribution with Node Differential Privacy. In: International Conference on Management of Data. San Francisco. pp. 123-138.

[22] Macwan, K.R., Patel, S.J. (2018) Node Differential Privacy in Social Graph Degree Publishing. Procedia Computer Science, 143: 786-793.

[23] Ullman, J., Sealfon, A. (2019) Efficiently Estimating Erdos-Renyi Graphs with Node Differential Privacy. Neural Information Processing Systems. Vancouver. pp. 3770-3780.

[24] Chen, S., Zhou, S. (2013) Recursive mechanism: towards node differential privacy and unrestricted joins. In: International Conference on Management of Data. Milan. pp. 653-664.

[25] Sabidussi, G. (1966) The centrality index of a graph. Psychometrika, 31(4): 581-603.

[26] Freeman, L.C. (1977) A set of measures of centrality based on betweenness. Sociometry, 1977: 35-41.

[27] Nissim, K., Raskhodnikova, S., Smith, A. (2007) Smooth Sensitivity and Sampling in Private Data Analysis. In: Symposium on the Theory of Computing. San Diego. pp. 75-84.

[28] Tang, L., Liu, H., Zhang, J., et al. (2008) Community Evolution in Dynamic Multi-mode Networks. In: ACM SIGKDD international conference on Knowledge discovery and data mining. Las Vegas. pp. 677-685.

[29] Brandes, U. (2001) A faster algorithm for betweenness centrality. Journal of mathematical sociology, 25(2): 163-177.

[30] Kourtellis, N., Alahakoon, T., Simha, R., et al. (2013) Identifying high betweenness centrality nodes in large social networks. Social Network Analysis and Mining, 3(4): 899-914.

[31] Wang, Y., Wu, X., Hu, D. (2016) Using Randomized Response for Differential Privacy Preserving Data Collection. EDBT/ICDT Workshops. Bordeaux. pp. 83-90.

[32] Individual contributions. (2016) Federal Election Commission. https://www.fec.gov/data/receipts/individual-contributions/?two_year_transaction_period=2016&min_date=01/01/2020&max_date=12/31/2016.

[33] Begg, C.B., Mazumdar, M. (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics, 1994: 1088-1101.

[34] Mcleod, A.I. (2020) Kendall: Kendall rank correlation and Mann-Kendall trend test. https://rdrr.io/cran/Kendall.

[35] Barabasi, A. (2009) Scale-Free Networks: A Decade and Beyond. Science, 325(5939): 412-413.