Journal of Computer and Knowledge Engineering, Vol. 6, No. 1, 2023. (37-46)
DOI: 10.22067/cke.2023.63240.0

Personalized Privacy Preserving Method for Social Networks Graph K-Anonymization*

Research Article

Hourie Mehrabiun[1] Behnaz Omoomi[2]

Abstract: Nowadays, with the development of social networks, the risk of disclosure of users' information has also increased, which has caused serious concerns among users. Accordingly, privacy preserving on social networks is a significant issue that has attracted much attention. Although there are various methods for preserving privacy on social networks, most of the existing methods are based on the universal approach that considers the same level of preservation for all users and only some of them consider individual personalized privacy requirements, which is very limited, and those are based on users' willing to share friends list and sensitive information with other users. This study focuses on a new scheme of personalized privacy preserving based on k-anonymity which can anonymize the social network graph based on the personalized privacy requirements of each individual. We develop a Modified Degree Privacy Level Sequence (MDPLS) Algorithm and execute experiments on two datasets. The results of the experiments show that in this new method of social network graph anonymization, when we consider the personalized privacy requirements, the costs of the anonymity process are reduced and data utility is improved in comparison with the situation where we only consider one level of privacy for all users, i.e., universal approach.

Keywords: Anonymous Social Network Graph, Personalized Privacy, Privacy Preserving, Social Network

1. Introduction

Social networks have made substantial development in recent years and are spreading quickly in different ways. Social networks are sources of worthwhile information that their publication is necessary and useful for data analysis.

Since this information contains sensitive and private data of many people, privacy preserving is one of the main concerns for using social networks. In order to preserve privacy, we have to publish an anonymous version of the network that differs from the original version. On the other hand, for the utility of analysis results of the anonymous version of the network, it should be similar as much as possible to the main network. Therefore, the significant problem is making balance between the security of the network and the loss of information in the released network [1].

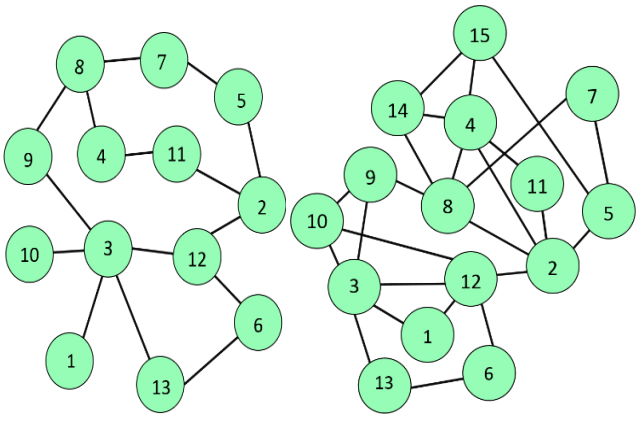
Three types of privacy breaches are defined in a social network [1, 2]:

Identity disclosure: Identity disclosure occurs when the identity of social network users is revealed. Identity disclosure causes the users' information and his relationship with other users be revealed too.

Sensitive link disclosure: Sensitive link disclosure occurs when an attacker finds out that a relationship exists between two users of the social network.

Sensitive attribute disclosure: Sensitive attribute disclosure occurs when an attacker gets accesses to the main and susceptible information of the social network users.

In this study, we consider the *identity disclosure*, which may occur when a social network is released. The goal is to protect the degrees of nodes in social networks. Usually, the node degree information of users is easy to access for adversaries. They can re-identify individuals and their relevant information in this way in the published social network graph. For example, if an adversary knows that one person has five friends in the network shown in Figure 1(a), he will directly find node 3 as the target person. In order to preserve the social network graph, against such degree attacks that caused vertex re-identification, Liu et al. [3] proposed a k-degree anonymity model. A graph $G(V,E)$ is k-degree anonymous if every node $u \in V$ has the same degree with at least k - 1 other nodes. Thus, an adversary cannot identify any vertex of the graph G with probability higher than $\frac{1}{k}$. Figure 1(b) shows a 5-degree anonymous version of the graph in Figure 1(a). An adversary cannot re-identify the target person in this graph with a probability higher than $\frac{1}{5}$.



(a) The original graph (b) A 5-degree anonymous graph

Figure 1. A degree anonymous graph without considering individual privacy requirements

In practice, different network users may not have the same

* Manuscript received: 2020 March 7, Revised, 2021April 19, Accepted, 2023 January 28.
1. Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran.
2. Corresponding author. Professor, Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran.
Email: bomoomi@iut.ac.ir

privacy preferences. In other words, people may have different levels of privacy requirements. Most of the proposed methods for privacy preserving do not consider the personalized privacy requirements of people. This may result in more anonymization costs and decrease the utility because of more information loss.

In the universal approach of social network anonymization, all nodes of the social network graph are anonymized based on a fixed amount. For instance, in the universal approach of k-anonymity, all nodes become k-anonymous while the users corresponding to some of these nodes may have no concern or have less caution about their privacy. Hence, k -anonymizing of the whole users just imposes an extra cost and leads to unnecessary changes in the graph structure. Users' more authority in determining their anonymity leads to better results based on the utility and cost measures.

Most studies on personalized privacy in social networks have considered people's privacy requirements into at most three categories.

Jiao et al. [4] introduced the personalized k-degree-l-diversity (PKDLD) anonymity model and specified three types of privacy attributes for various individuals. Lan et al. [5] divided entities into sensitive and non-sensitive and proposed the *k*-neighborhood anonymous method. Yuan et al. [6] defined three levels of protection requirements for users and used label generalization protection and structure protection techniques on an unweighted graph with labels on both nodes and edges.

We propose in this study a new scheme of personalized privacy preserving based on the k-anonymity. First, each user is asked to determine his own privacy level. Then we anonymize the social network graph based on the given level requirement of each individual. These levels can be more than three levels and each individual in the social network graph has his own privacy level requirement.

2. Review of related studies

A variety of works for preserving privacy in microdata (tabular data) had been done. Samarati et al. [7] defined k-anonymity for a data table. A table is *k*-anonymous when its data cannot match with smaller than k individuals. They introduced a generalized table and minimal generalization of a table with respect to *k*-anonymity requirements and proposed the Minimal Generalization (MinGen) algorithm that for the given table returns the preferred minimal generalization. Truta et al. [8] focused on the *k*-anonymity approach presented by Sweeney et al. and Samarati [9, 10], and they showed that the *k*-anonymity model protects against identity disclosure but it fails to protect against attribute disclosure. They proposed a new method to preserve the privacy property called p-sensitive k-anonymity model that preserves both identity and attribute disclosure. Aggarwal et al. [11] considered a new framework for privacy preserving data mining that the privacy requirements of records are different. They proposed a new method based on the condensation approach [12]. Their experiments on some real data sets showed the effectiveness of this new method. Xiao et al. [13] proposed an approach for preserving the privacy in publication of sensitive data that is based on personalized anonymity. They developed a new generalization method that satisfies individual requirements, with the minimum amount of necessary generalization and the maximum amount of information in the microdata. Another approach to personalized privacy preservation of microdata was proposed by Shen et al. [14] that was based on the decision tree classification algorithm. Moreover, Xu et al. [15] presented a personalized k-anonymity method that is based on generalizing the quasi-attribute values by hierarchy to anonymize microdata. Ford et al. [16] introduced the p-sensitive k-anonymity model which is a combination of the k-anonymity model in [17] and the p-sensitive k-anonymity model for microdata in [8]. They also integrated available algorithms for the p-sensitive k-anonymity for microdata [18] and the k-anonymity for social networks [17] into a new algorithm called "SaNGreeA PK".

Many other methods of privacy preserving and anonymizing algorithms on microdata have been proposed based on k-anonymity such as l-diversity [19], t-closeness [20], and [21, 22, 23, 24].

With the development of social networks and the expansion of their use, the methods of privacy protection in social networks have been rapidly developed too. Hay et al. [25] proposed a new technique for social network anonymization against the re-identification attack. This method that is based on perturbing the network does not modify nodes but makes some edge deletions and edge insertion in a random sequence. They showed that this technique considerably reduces the privacy threats. Based on this, Liu et al. [3] studied the k-degree anonymity to preserve the social network against identity disclosure. They assumed that an attacker has some background knowledge of the degree. They first considered only edge addition (or similarly edge deletion), and also extended their proposed method such that both operations of edge addition and deletion be simultaneously allowed to modify the input graph. Zhou et al. [26] assumed that the attacker has some background knowledge about the neighborhood of some individuals and their relationships, so they identified another type of social network privacy attack, i.e., neighborhood attack. For social network anonymization, they focused on k-anonymity and handled only 1-neighborhood. Campan et al. [27] proposed a new method for anonymizing social network data. In this approach that is based on edge generalization a greedy algorithm, "SaNGreeA" is presented. They also defined a measure to quantify information loss. Zou et al. [28] considered the identity disclosure problem for social networks. To preserve the network against structural attacks, they proposed the k-automorphism and "KM" algorithm. They also extended this algorithm to the dynamic release of the networks. Zhou et al. [29] modeled neighborhood attacks and extended the k-anonymity and l-diversity models to preserve the privacy of the social network. Yuan et al. [30] proposed the k-degree-l-diversity "KDLD" (k-degree-l-diversity) model to prevent re-identification of individuals in social networks and their sensitive attributes. Ninggal et al. [31] proposed the "utility-aware social network graph anonymization" to anonymize the relationship information of the social network graph and protect the identity of individuals. Their method is based on preserving the structural properties of the social network graph. Macwan et

al. [32] proposed a new clustering approach and presented an improved k-degree anonymity model that preserves the privacy of individuals and low utility loss.

In the real world, different users of a social network may have different levels of privacy requirements. Most of the proposed methods of privacy preserving do not consider various requirements of privacy level for individuals. This may cause more anonymization cost and decrease the utility because of more information loss. Most studies on personalized privacy in social networks classified individuals' privacy requirements into at most three categories.

Yuan et al. [6] designed a framework for privacy protection on labeled social networks. They defined three levels of privacy based on the attacker's background knowledge. In level 1 the attacker only knows the users' labels. In level 2 in addition to the labels of nodes, their degree is also known for the attacker and in level 3 the attacker knows the labels of the adjacent edges to the nodes too. They proposed methods to preserve the privacy of users according to the desired framework in each level of protection. Lan et al. [5] used the k-neighborhood anonymity approach based on different privacy protection levels. In this method, entities are divided into sensitive entities with the privacy protection requests and non-sensitive entities that do not need to protect privacy. They wanted to preserve the sensitive entities from rediscovering. They introduced "KNAP'' algorithm that gets the social network graph G and parameter k, and returns the k-neighborhood anonymous publication of the graph G. Babu et al. [33] proposed Compute Sensitivity Index (D, L) to compute the sensitivity index of each node that indicates the importance of it in a network. They presented a generalization approach to anonymizing users in social networks that focused on the importance of users. Jiao et al. [4] proposed the "personalized k-degree l-diversity (PKDLD) anonymity model" with a focus on protecting the degree of nodes and one sensitive label of them. They assumed that an adversary with knowledge about some users' degree wants to re-identify an individual. They also considered the privacy requirements of users. They classified the privacy requirements into three levels, H (high), M (middle), and L (low), according to the users' willingness to the accessibility of other people to their friends' list and sensitive attributes. They also designed and implemented PKPALDP algorithm to graph anonymization and showed that their algorithm has better results in comparison with the existing approaches.

In this paper, we propose a new method for anonymizing a social network graph based on the different privacy level requirements of each individual. This type of personalized privacy method is based on k-degree anonymity [3] and can be more efficient than the universal k-anonymity approach of privacy preserving methods because a certain level of privacy is considered for each individual that is determined by the user. Therefore, it reduces the cost of anonymizing and improves the utility of data.

The structure of the paper is as follows. In Section 3 we describe the problem and present some necessary definitions. Then, we briefly review our proposed method. In Section 4, we present Algorithm 1 (MDPLS) and Algorithm 2 (PKLADP) to solve the given problem. In Section 5, we present the results of experiments on two datasets and we discuss and analyze the results of the experiments. The Conclusion of the paper and some future works are given in section 6.

3. The statement of the problem

This study focuses on the personalized privacy preserving problem for a social network graph. We consider a suitable framework such that each user is allowed to determine his own special privacy level. We assume that an attacker has access to some background knowledge about the degree of some nodes and wants to re-identify a known individual in the published social network graph. Our goal is to preserve the privacy of each user by adding certain edges/nodes to transform a social network graph into a personalized degree anonymous graph.

More precisely the problem that we consider is as follows. The input of the problem is a simple graph G corresponding to a social network and a set $L = \{l_1, l_2, \cdots, l_{|V|}\}$ that l_i is the privacy level of the corresponding user with node i in the vertex set V (for convenience we say privacy level of the node i) which is determined by the user. We need some graph modification operations on G to construct an anonymous graph G^* that is structurally similar to the original graph G as much as possible and for each node v with privacy level l_v, the probability that an attacker re-identifies v is at most $\frac{1}{l_v}$.

Definition 1. A *social network graph* is a four tuple $G = (V, E, L, \gamma)$, where V is the set of nodes, E is the set of unordered pairs of nodes called edges, L is the set of privacy level of nodes, and $\gamma : V \rightarrow L$ is the privacy level function that maps each node to its privacy level requirement l_v in L, that is, the privacy level of the node v.

In the next definition, we assign a triple to each node and we arrange nodes in descending order of degree.

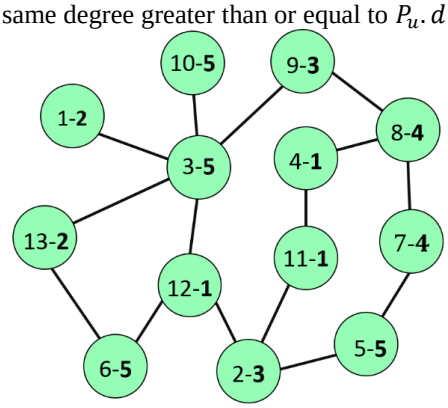
Definition 2. Given a social network $G = (V, E, L, \gamma)$ with $|V| = n$, the Degree Privacy Level Sequence (DPLS) P for G is a sequence $[P[1], P[2], \dots, P[n]]$, where for each $i, 1 \leq i \leq n$, $P[i]$ is a three tuple $(id, P_i.d, P_i.l)$ that id identifies the node, $P_i.d$ is the degree of the node i and $P_i.l$ is the privacy level of it such that $P_1.d \geq P_2.d \geq \dots \geq P_n.d$, and whenever $P_i.d = P_j.d$, then $l_i \geq l_j$. The value of $|V|$ determines the size of the sequence.

In Figure 2(a), for example, the degree privacy level sequence is as follows (the numbers corresponding to the privacy level of nodes are shown bold in the figure).

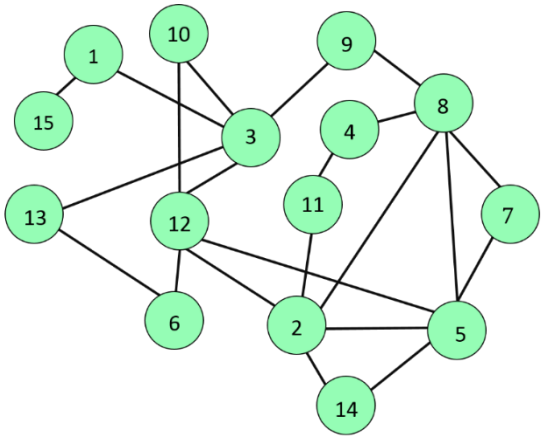
$P = [(3,5,5), (8,3,4), (2,3,3), (12,3,1), (5,2,5), (6,2,5), (7,2,4), (9,2,3), (13,2,2), (4,2,1), (11,2,1), (10,1,5), (1,1,2)]$

Definition 3. A sequence P' is called Modified Degree Privacy Level Sequence (MDPLS) of degree privacy level sequence P, if P' can be divided into subsequences $[P'[1], \dots, P'[i_1]], [P'[i_1 + 1], \dots, P'[i_2]], \dots, [P'[i_m + 1], \dots, P'[j]]$, called anonymous classes such that every node u belongs to an anonymous class of P' say P'_x=$[P'[i_x]$,…,

$P'[\, i_{x+1}]]$, where $i_{x+1} - i_x + 1 \geq l_u$ and all elements in P'_x have the same degree greater than or equal to $P_u.d$.



(a) The original graph with the privacy level of each node



(b) The personalized degree anonymous graph

Figure 2. A degree anonymous graph considering the individuals' privacy requirements

We propose a two-step approach to the personalized graph anonymization problem as follows.

1. First, for the given degree privacy level sequence P of the social network graph $G = (V, E, L, \gamma)$, we construct a modified degree privacy level sequence P' with the goal of minimizing $L(P', P) = \sum_{u \in V} |P'_u.d - P_u.d|$, where $P_u.d$ is the degree of node u in P. In fact, $L(P', P)$ is the degree difference between these two degree sequences. Clearly, the smaller degree difference is equivalent to the more structural similarity between the published and the original graph.
2. Based on the input graph $G = (V, E)$ and the modified degree privacy level sequence P', we use PKLADP algorithm that was proposed in [4] with minor changes that are required for our purpose, to construct a new graph G^* = (V^*, E^*) that realizes the degree sequence P^* and $E \subseteq E^*$ and satisfies the graph privacy requirement for each individual. The degree sequence P^* of G^* is a personalized anonymous degree sequence and has all the elements in P'.

The personalized degree anonymous graph G^* is constructed from the original graph G by adding some noise nodes and edges.

4. Proposed algorithm

In this section, we propose our algorithm to generate a modified degree privacy level sequence. Then, we use PKLADP algorithm [4] to personalized degree anonymous graph construction.

4.1. Modified degree privacy level sequence generation

Given a degree sequence P of the original social network graph G, in order to generate a modified degree privacy level sequence P', we divide the degree sequence P into anonymous classes such that every node v in each anonymous class is at least l_v anonymous. We assume that the size of P equals n and for simplicity of notation we use v_j to denote j-th node in P.

Algorithm 1. Modified Degree Privacy Level Sequence (MDPLS)

Require: The degree privacy level sequence P of the original social network graph G.
Ensure: The modified degree privacy level sequence P'.
1: $i \leftarrow 1, num \leftarrow n$.
2. **while** ($num \neq 0$)
3: choose the first unclassified node u, in the sequence P
4: $l_{current} \leftarrow P_u.l$
5: **if** ($l_{current} \leq num$)
6: $l_{max} \leftarrow$maximum privacy level of the next $l_{current}$ nodes from u
7: **while** ($l_{max} > l_{current}$ *and* $l_{max} \leq num$)
8: $l_{current} \leftarrow l_{max}$
9: $l_{max} \leftarrow$ maximum privacy level of the next $l_{current}$ nodes from u
10: **end while**
11: **if** ($l_{max} \leq num$)
12: consider anonymous class C_i of size l_{max} from node u
13: update the degree sequence of nodes in the class C_i
14: $i \leftarrow i + 1,\ num \leftarrow num - |C_i|$
15: **else**
16: $l_r \leftarrow l_{max} - num$
17: **if** (there is node v_j before u in the sequence P such that $P_{v_j}.l \geq l_r$)
18: merge the unclassified nodes into the anonymous class that v_j belongs to it and update degree sequence of its nodes
19: $num \leftarrow 0$
20: **else**
21: $j \leftarrow n - l_{max} + 1$
22: merge the $l_{max} - 1$ nodes at the end of sequence into anonymous class that v_j belongs to it and update degree sequence of its nodes
23: $num \leftarrow 0$
24: **else**
25: $l_{max} \leftarrow$maximum privacy level of unclassified nodes
26: $l_{max} \leftarrow max\{l_{max}, l_{current}\}$
27: repeat lines 16-23
28: **end while**

In Algorithm 1, the given degree privacy level sequence P is anonymized as follows. We select the first un-anonymous node u, the node that doesn't belong to any anonymous class, of the sequence and then construct the anonymous class in two steps. First, we construct the initial anonymous class, starting from node u to l_u next nodes, and then we extend the initial class by merging the next nodes in the sequence into it until the maximum privacy level of the nodes in the current anonymous class is equal to its size. Now the degree sequence of nodes in the anonymous class is updated by

increasing the degree of all nodes in it to the maximum degree in this subsequence.

Each time a new anonymous class is constructed the number of unclassified nodes is reduced by the size of the class and we repeat this process until all nodes in the sequence P are anonymized.

Note that it is possible that the remaining unclassified nodes be strictly less than the number of required nodes to create or extend an anonymous class. In these cases, we merge the unclassified nodes into a proper existing anonymous class in one of the following ways.

Let num and l_{max} be the number of unclassified nodes and required nodes to create or extend an anonymous class, respectively. If an anonymous node v_j whose privacy level is at least $(l_{max} - num)$ exists, by merging the unclassified nodes into the class C that $v \in C$, and then updating the degree sequence of nodes in the extended class, the remaining nodes can be anonymized. Otherwise, we consider the existing anonymous class that l_{max}-th node from the end of the sequence belongs to it as the last anonymous class, and we merge the nodes after it into this class and then we update the degree sequence of nodes in it.

Since the nodes in the degree privacy level sequence P are sorted based on descending order of degree and privacy level, using the above method, nodes with similar degree and similar privacy levels are classified together. For example, by applying the above algorithm on degree sequence of graph Figure 2(a), we have:

$P' =$
$[(3,5,5),(8,5,4),(2,5,3),(12,5,1),(5,5,5),(6,2,5),$
$(7,2,4),(9,2,3),(13,2,2),(4,2,1),(11,2,1),(10,2,5),$
$(1,2,2)\ ]$

Algorithm 1 returns the modified degree privacy level sequence P' in time $O(n)$ for a given degree privacy level sequence P of size n. The procedure scans the degree sequence P at most twice; once for constructing anonymous classes and at most once again to handle the special situation for the last anonymous class in the way that was explained. Hence, the time complexity of providing the MDPLS is $O(n)$.

4.2. Graph construction

Now for the given graph G with the degree privacy level sequence P, we have a modified degree privacy level sequence $P^{'}$, that is obtained from Algorithm 1 and we want to construct a new graph G^* based on $P^{'}$ with the maximum structural similarity to the original graph G, in other words, by performing minimal changes on the original graph G. We call such graph G^*, *personalized degree anonymous graph*.

As we saw in Algorithm 1, we want all nodes that had been classified in the same class to have an identical degree. We consider the maximum degree of nodes in each class C_i, as the target degree of it and then, we increase the degree of all nodes in each anonymous class to this amount.

In the next definition, we assign to each node a value that indicates the required degree of each node in P to reach the desired degree in P'.

Definition 4. Given the degree privacy level sequence P and modified degree privacy level sequence P' as the output of Algorithm 1, the excess of a node v in P' with respect to P, denoted by $\text{ex}.d_v$ is:

$$\text{ex}.d_v = P_{v'}.d - P_v.d \quad (1)$$

We use PKLADP algorithm [4] with minor changes for constructing a personalized anonymous graph of the given graph G and its modified degree privacy level sequence. We consider a set V_{in} that includes all nodes with nonzero degree excess. In fact, V_{in} is the set of all nodes whose degree in the P is smaller than their target degree, i.e., their degree in the P'.

Then, for each node v in V_{in}, if there is another node u in V_{in}, such that the distance between u and v in the original graph G is two, we connect u and v. By such operations on G, the degrees of both u and v increase by one which leads to a decrease in the degree excess of them by one. On the other hand, the distance between u and v decreases to one and therefore the length of the shortest paths between u and v changes by 1. If the degree excess of any node u or v is achieved to zero, we remove it from the set V_{in}.

Algorithm 2. Personalized Degree Anonymous Graph Construction (PKLADP)

Require: The original graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, degree privacy level sequence $\boldsymbol{P}$ of $\boldsymbol{G}$, and modified degree privacy level sequence $\boldsymbol{P'}$
Ensure: Personalized degree anonymous graph
1: create the set $\boldsymbol{V_{in}}$ including all nodes that degree excess of them is nonzero
2: if $\boldsymbol{V_{in}}$ is not empty
3: for all nodes $\boldsymbol{v}$ in $\boldsymbol{V_{in}}$
4: if exists some node $\boldsymbol{u}$ in $\boldsymbol{V_{in}}$ such that $\boldsymbol{d(u,v) = 2}$
5: link $(\boldsymbol{u}, \boldsymbol{v})$
6: increase the degree of $\boldsymbol{u}$ and $\boldsymbol{v}$ by one
7: end if
8: if $\boldsymbol{ex.d_u = 0}$
9: remove $\boldsymbol{u}$ from $\boldsymbol{V_{in}}$
10: end if
11: if $\boldsymbol{ex.d_v = 0}$
12: remove $\boldsymbol{v}$ from $\boldsymbol{V_{in}}$
13: end if
14: end for
15: end if
16: while $\boldsymbol{V_{in}}$ is not empty
17: for all node $\boldsymbol{v}$ in $\boldsymbol{V_{in}}$
18: add a noise node $\boldsymbol{x}$ which $\boldsymbol{l_x = 1}$ to graph and connect it to $\boldsymbol{v}$
19: increase the degree of $\boldsymbol{v}$ by one
20: if there is another node $\boldsymbol{u}$ in $\boldsymbol{V_{in}}$ such that $\boldsymbol{d(u,v) = 1\ or\ d(u,v) = 2}$
21: connect $\boldsymbol{x}$ to $\boldsymbol{u}$ and increase the degree of $\boldsymbol{u}$ by one
22: end if
23: if $\boldsymbol{e_x.d_v} = 0$
24: remove $\boldsymbol{v}$ from $\boldsymbol{V_{in}}$
25: end if
26: if $\boldsymbol{e_x.d_u} = 0$
27: remove $\boldsymbol{u}$ from $\boldsymbol{V_{in}}$
28: end if
29: end for
30: end while

If V_{in} is not still empty, we add a noise node x that we set $l_x = 1$, for each node v in V_{in} and connect it to v. The degree of v increases by 1. If its degree excess is achieved to zero, we remove it from V_{in}. If there is another node u in V_{in} such that $d(u,v) = 1$ or $d(u,v) = 2$, (d denotes distance in the original graph) we also connect x to u. The degree of u increases by 1 and same as before if the degree excess of them is achieved to zero we remove them from V_{in}.

Note that we are just allowed to add noise edges and nodes in the graph constructing process, as described above.

The time complexity of PKLADP algorithm is $O(n^2)$[4].

5. Experiments

In this section, using two datasets we evaluate the performance of the proposed personalized graph anonymization method and we compare it with the universal approach.

The first dataset is the email network of University Rovira i Virgili (URV) (http://konect.cc/networks/arenas-email/) with 1133 nodes and 5451 edges. This graph shows email interchanges between members of the university. Nodes are users and each edge represents that at least one email was sent.

The second dataset contains information about the power grid of the Western States of the United States of America (http://konect.cc/networks/opsahl-powergrid) with 4941 nodes and 6594 edges. An edge represents a power supply line. A node is either a generator, a transformator, or a substation.

5.1. Results and analysis

In this section, we will evaluate the effectiveness of our algorithm. Our main objective is to achieve higher data utility and less information loss in the personalized setting of anonymization, which allows users to have different levels of privacy, compared to the universal approach, which considers a fixed level of privacy for all users. We perform our experiments on the two above datasets.

In the personalized setting, we consider different privacy levels $l = \{5,10,15,20,25,30\}$ for users. For each value l, we attribute the privacy level between 1 and l randomly to users. In practice we allow users to determine their privacy level from a range of privacy levels and we compare the results of this approach with the results of universal approach that all users have the same privacy level l.

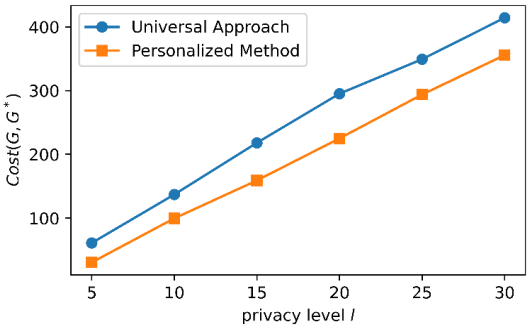
5.1.1. Anonymous cost

To measure the information loss, we compute two types of anonymous cost, $L(P', P)$ and $Cost(G, G^*)$ [4]. $L(P', P)$ computes the total degree increase of all nodes from degree privacy level sequence P of the original graph G to the modified degree privacy level sequence P' by using Algorithm 1, that is

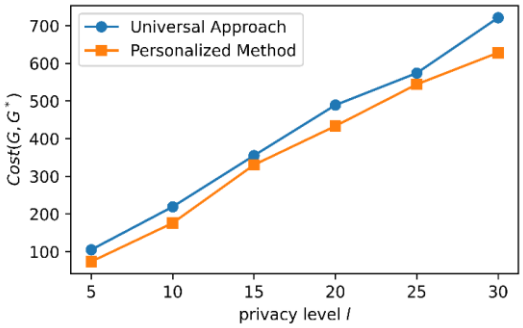
$$L(P', P) = \sum_{\forall u \in V} |P'_u.d - P_u.d|.$$

$Cost(G, G^*)$ is the sum of the number of nodes and edges that were added from the original graph $G = (V, E, L, \gamma)$ to the published graph $G^* = (V^*, E^*, L^*, \gamma^*)$ with $V \subseteq V^*$ and $E \subseteq E^*$ that is:

$$Cost(G, G^*) = (|E^*| - |E|) + (|V^*| - |V|).$$

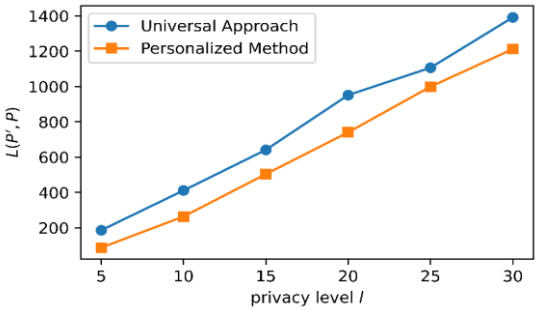


(a) URV

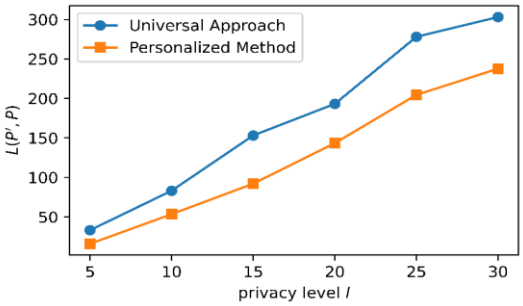


(b) US Power grid

Figure 3. URV dataset and US power grid dataset: Cost(G, G*)



(a) URV



(b) US Power grid

Figure 4. URV dataset and US power grid dataset: $L(P', P)$

Figures 3 and 4 show the above measures of cost in the two approaches. One is the universal approach that considers identical values l for privacy level of all users. The other is the personalized method that attributes different values of privacy level between 1 and l to each user. In the personalized setting, we repeated the experiment 20 times for each privacy level l and we reported the average of

$Cost(G, G^*)$ and $L(P', P)$ over 20 experiments. As we see $Cost(G, G^*)$ and $L(P', P)$ both are reduced in the personalized privacy method compared to the universal approach. Accordingly, based on the anonymization cost, the personalized method is better than the other.

5.1.2. Utility

To measure the quality of the published anonymous graph and its similarity to the original graph, we define two new measures $APEPL(G, G^*)$ (Average percentage error of shortest path length) and $APECC(G, G^*)$ (The average percentage error of clustering coefficient) as follows.

$$APEPL(G, G^*) = \frac{\sum_{u,v \in V} \frac{lp_G(u,v) - lp_{G^*}(u,v)}{lp_G(u,v)} * 100}{n(n-1)}$$

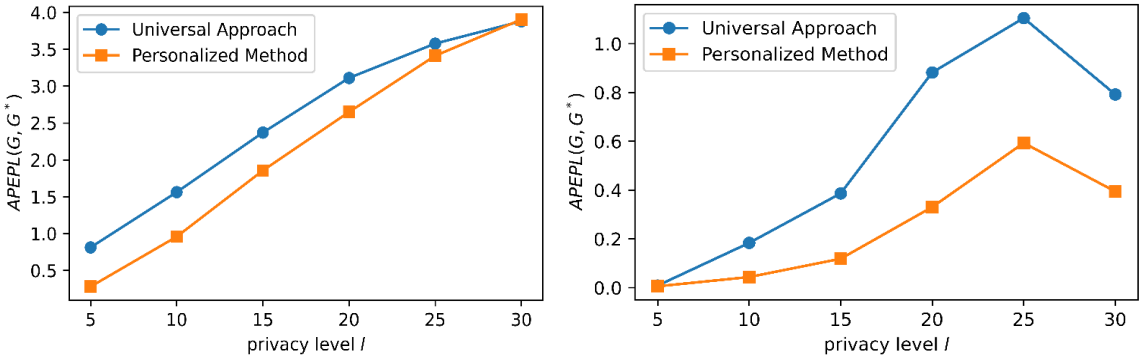
where $lp_G(u, v)$ denotes the shortest path length between nodes u and v in the graph G and n is the number of its nodes.

$$APECC(G, G^*) = \frac{\sum_{v \in V \; s.t \; cc_v(G)!=0} \frac{cc_v(G) - cc_v(G^*)}{cc_v(G)} * 100}{|V_c|}$$

where $cc_v(G)$ is the clustering coefficient of node v in the graph G and $|V_c|$ is the number of its nodes whose clustering coefficient is non-zero.

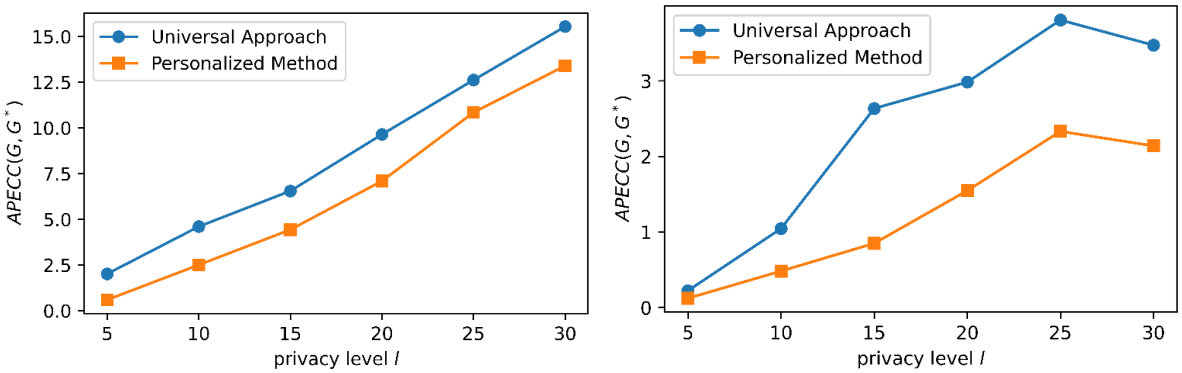
Figures 5 and 6 show $APEPL(G, G^*)$ and $APECC(G, G^*)$ in which G^* is the personalized degree anonymous graph based on two settings of anonymization. In the personalized method, the experiments are repeated 20 times and the average of results over 20 experiments are reported. As we see, the average percentage error of path length of G and G^* is almost perfectly less in the case of using personalized settings and the average percentage error of clustering coefficient of G and G^* is also less in the case of using personalized settings.

According to these comparisons, the utility of the anonymous social network graph in the proposed method increases by considering personalized privacy level for each individual.



(a) URV (b) US Power grid

Figure 5. URV dataset and US power grid dataset: APEPL Cost(G, G*)



(a) URV (b) US Power grid

Figure 6. URV dataset and US power grid dataset: $APECC(G, G^*)$

6. Conclusion and future studies

We proposed in this study a new method for personalized anonymization of social network graphs based on the k-anonymity model. We defined a new framework in which all users are allowed to set their own privacy level requirements. We proposed an algorithm to modified degree sequence of social network graph such that satisfies personalized anonymity for each node and then based on the obtained degree sequence and using PKLADP algorithm, constructed a personalized degree anonymous graph. We applied our proposed algorithms to two data sets and showed that personalized method improves the data utility and reduces the cost of anonymization compared to the case where only one constant privacy level is considered for all individuals.

There are several future works that could be explored. One is extending the algorithms and analysis using other privacy preserving methods in social networks. The other is developing a personalized privacy preserving method for a social network where users' privacy level requirements constantly change. Another future study can consider social network as a directed graph and provide personalized privacy preserving methods in a framework where each node corresponding to a user could have two specific privacy levels for its in-degree and out-degree. Moreover, it will be interesting to deal with other types of privacy breaches in social networks, such as link disclosure. Future studies acn provide personalized privacy preserving methods for this purpose.

7. References

[1] A. Singh, D. Bansal, S. Sofat, "Privacy preserving techniques in social networks data publishing-a review", *International Journal of Computer Applications*, vol. 87, no. 15,2014

[2] E. Zheleva, L. Getoor, "Preserving the privacy of sensitive relationships in graph data," in: *International Workshop on Privacy, Security, and Trust in KDD*, pp. 153-171: Springer, 2007.

[3] K. Liu, E. Terzi, "Towards identity anonymization on graphs*," in: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pp. 93-106: ACM, 2008.

[4] J. Jiao, P. Liu, X. Li, "A personalized privacy preserving method for publishing social network data," *in: International Conference on Theory and Applications of Models of Computation*, pp. 141-157: Springer, 2014.

[5] L. Lan, H. Jin, Y. Lu, "Personalized anonymity in social networks data publication," *in: 2011 IEEE International Conference on Computer Science and Automation Engineering*,
vol. 1, pp. 479-482: IEEE, 2011.

[6] M. Yuan, L. Chen, P. S. Yu, "Personalized privacy protection in social networks*," Proceedings of the VLDB Endowment*, vol. 4 no. 2, pp. 141-150, 2010.

[7] P. Samarati, L. Sweeney, "Generalizing data to provide anonymity when disclosing information," *in: PODS*, vol. 98, no. 10.1145 pp. 275487-275508, 1998

[8] T. M. Truta, B. Vinay, "Privacy protection: p-sensitive k-anonymity property," *in: 22nd International Conference on Data Engineering Workshops (ICDEW'06),* pp. 94-94: IEEE, 2006.

[9] P. Samarati, "Protecting respondents identities in microdata release," *IEEE transactions on Knowledge and Data Engineering*, vol. 13, no. 6 pp. 1010-1027, 2001.

[10] L. Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 571-588, 2002.

[11] C. C. Aggarwal, P. S. Yu, "On variable constraints in privacy preserving data mining," *in: Proceedings of the 2005 SIAM International Conference on Data Mining, SIAM*, pp. 115-125, 2005.

[12] C. C. Aggarwal, S. Y. Philip, "A condensation approach to privacy preserving data mining," *in: International Conference on Extending Database Technology, Springer*, pp. 183-199, 2004.

[13] X. Xiao, Y. Tao, "Personalized privacy preservation," *in: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, ACM*, pp. 229-240, 2006.

[14] Y. Shen, H. Shao, Y. Li, "Research on the personalized privacy preserving distributed data mining," *in: 2009 Second International Conference on Future Information Technology and Management Engineering, IEEE*, pp. 436-439, 2009

[15] Y. Xu, X. Qin, Z. Yang, Y. Yang, K. Li, "A personalized k-anonymity privacy preserving method," *Journal of Information & Computational Science*, vol. 10, no. 1, 2013 pp. 139-155.

[16] R. Ford, T. M. Truta, A. Campan, "P-sensitive k-anonymity for social networks," *DMIN*, vol. 9, pp. 403-409, 2009

[17] A. Campan, T. M. Truta, "A clustering approach for data and structural anonymity in social networks," *in: International Workshop on Privacy, Security, and Trust in KDD*, 2008.

[18] A. Campan, T. M. Truta, j. Miller, R. Sinca, "A clustering approach for achieving data privacy," *in: International Conference on Data Mining DMIN*, 2007.

[19] A. Machanavajjhala, J. Gehrke, D. Kifer, M. Venkitasubramaniam, "ldiversity: Privacy beyond k-anonymity," *in: 22nd International Conference on Data Engineering (ICDE'06), IEEE*, pp. 24-24, 2006.

[20] N. Li, T. Li, S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," *in: 2007 IEEE 23rd International Conference on Data Engineering, IEEE*, pp. 106-115, 2007.

[21] G. Aggarwal, R. Panigrahy, T. Feder, D. Thomas, K. Kenthapadi,S. Khuller, A. Zhu, "Achieving anonymity via clustering," *ACM Transactions on Algorithms (TALG)*, vol. 6, no. 3, pp. 1-19, 2010.

[22] T. Li, N. Li, J. Zhang, I. Molloy, Slicing: "A new approach for privacy preserving data publishing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 3, pp. 561-574, 2010.

[23] X. Xiao, Y. Tao, "Anatomy: Simple and effective privacy preservation," *in: Proceedings of the 32nd International Conference on Very Large Data Bases*,

VLDB Endowment, pp. 139-150, 2006.

[24] F. Amiri, N. Yazdani, A. Shakery, A. H. Chinaei, "Hierarchical anonymization algorithms against background knowledge attack in data releasing," *Knowledge-Based Systems*, vol. 101, pp. 71-89, 2016.

[25] M. Hay, G. Miklau, D. Jensen, P. Weis, S. Srivastava, "Anonymizing social networks," *Computer science department faculty publication series*, pp.1-180, 2007.

[26] B. Zhou, J. Pei, "Preserving privacy in social networks against neighborhood attacks," *in: 2008 IEEE 24th International Conference on Data Engineering*, pp. 506-515, 2008.

[27] A. Campan, T. M. Truta, "Data and structural k-anonymity in social networks", *in: International Workshop on Privacy, Security, and Trust in KDD, Springer*, pp. 33-54, 2009.

[28] L. Zou, L. Chen, M. T. Ozsu, "K-automorphism: A general framework for ¨privacy preserving network publication," *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 946-957, 2009.

[29] B. Zhou, J. Pei, "The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks," *Knowledge and Information Systems*, vol. 28, no. 1, pp. 47-77, 2009.

[30] M. Yuan, L. Chen, S. Y. Philip, T. Yu, "Protecting sensitive labels in social network data anonymization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 3, pp. 633-647, 2011.

[31] M. I. H. Ninggal, J. H. Abawajy, "Utility-aware social network graph anonymization," *Journal of Network and Computer Applications*, vol. 56, pp. 137-148, 2015.

[32] K. R. Macwan, S. J. Patel, "k-degree anonymity model for social network data publishing," *Advances in Electrical and Computer Engineering*, vol. 17, no. 4, pp. 117-125, 2017

[33] K. S. Babu, S. K. Jena, J. Hota, B. Moharana, "Anonymizing social networks: A generalization approach," *Computers & Electrical Engineering*, vol. 39, no. 7, pp. 1947-1961, 2013.