Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7187528, 11 pages
https://doi.org/10.1155/2022/7187528

Research Article

A Personalized (α, β, l, k)-Anonymity Model of Social Network for Protecting Privacy

Xiangmin Ren[1] **and Dexun Jiang**[2]

[1]*College of Computer Science and Technology, Taizhou University, Taizhou, Jiangsu Province, China*
[2]*School of Information Engineering, Harbin University, Harbin, Heilongjiang Province, China*

Correspondence should be addressed to Xiangmin Ren; a01905@tzu.edu.cn

Received 6 January 2022; Revised 18 February 2022; Accepted 7 March 2022; Published 20 April 2022

Academic Editor: Deepak Kumar Jain

By mining the data published on social network, we can discover the hidden value of information including the privacy of individuals and organizations. Protecting privacy of individuals and organizations on social network has become the focus of more and more researchers. Based on the actual privacy protection need of edge sensitive attribute and vertexes sensitive attribute, we propose a new personalized (α, β, l, k)-anonymity technology of privacy preserving to reduce distortion extent of the data in the privacy processing of data of social network. Experimental results of personalized (α, β, l, k)-anonymity algorithm show that d-neighborhood attack of graph, background knowledge attack, and homogeneity attack can be prevented effectively by using anonymous vertexes and edges, as well as the influence matrix based on background knowledge. The diversity of vertex sensitive attribute can be achieved. Personalized protecting privacy requirements can be met by using such parameter as α, β, l, k.

1. Introduction

Data publishing of social network is very important for scientific research, commercial purpose, countries, and so on, but social network data includes privacy information and sensitive relations, which can be leaked by publishing directly. How to protect individual privacy, and make the publishing data or graph useful at the same time, has become very important problem of social network data publishing. One of the most important principle is that individual can decide his own privacy whether be published or not, that is to say individual has different privacy protection needs.

Anonymity techniques for data publishing have been used in the relational data for a long time, and make great progress in relational database area, including k-anonymity, l-diversity, generalization, and so forth [1, 2]. Can we apply the same anonymous techniques that apply to relational data to social networks? Social network data contains more information than relational data because network data contains vertexes (nodes), edges, relationships between nodes, and various metric features of the graph. Some researchers want to use these technologies on data publishing of social network [3]. So the structure and evaluation of social network method were proposed in paper [4], and the categories of attack in social network can be found in paper [5]. Graph modification [6], graph partitioning [7], graph isomorphism [8], clustering [9], attribute generalization [10], and so on were applied to data publishing of social network, and then more and more anonymous technologies of social networks appear in many academic papers.

Actually, a graph structure is necessary to represent the network vertexes relations rather than a two dimensional representation in relational database [11], degree denotes the relationship between two vertexes, high degrees mean the relationships are more closer among the vertexes, and there are only a small part of vertexes which degrees are high, degrees of most of vertexes are low in big social network. So a limited fraction of vertexes with high degrees bring a lot of data loss and computation cost when using unified anonymity methods and the same privacy protection level [12].

Personalized privacy protection based on data table was proposed firstly by Xiao and Tao in 2006 [13]. They used individual guarding node to set level of self-sensitive attribute and did not set the same anonymity level for all individuals, but rather anonymity according to setting guarding node.

Ever since then, more and more researcher paid more attention to personalized anonymity of data publishing and made modest progress. During the research process of social network privacy protection, because data of social network is more complex than traditional data table, most of social network research used unified anonymity methods and the same privacy protection level. For example, user can create their basic information, Web albums, Web logs, the lists of friends, and so on. But Facebook, Twitter, Wechat, and voov meeting, they were able to decide those information whether can be accessed and viewed by others according to their own privacy level, consequently achieved purpose of preserving privacy to some extent.

The data in social network is more complex than two-dimensional data table in relational database. Privacy protection in social network can be summarized as vertex protection, edge protection, and sensitive attribute protection. Vertex protection is to prevent an attacker from identifying a vertex in an anonymous publishing graph with a high probability. Edge protection is to prevent an attacker from identifying an edge in an anonymous publishing graph with a high probability. Attribute protection is to prevent the attacker from getting vertexes or sensitive attributes of edges with a high probability. We cannot use anonymity methods and technologies, which used into traditional two dimensional data table, into social network directly, and users have personalized protecting privacy requirements (vertex protection, edge protection, and sensitive attribute protection) in the real social network such as the users of Facebook, Twitter, and Wechat, so it has the very high research value that personalized privacy protection methods are used into social network data publishing [14].

2. Problem Definition

2.1. Related Concepts

Definition 1. k-Anonymity. $\mathrm{RT}(A_1, \cdots, A_n)$ is a table and Q I_{RT} is quasi-identifier in RT. RT is said to satisfy k-anonymity if and only if each sequence of values in $\mathrm{RT}[\mathrm{QI}_{\mathrm{RT}}]$ emerge k occurrences at least in $\mathrm{RT}[\mathrm{QI}_{\mathrm{RT}}]$ [15].

Table 1 is said to satisfy k-anonymity, $\mathrm{QI}_{\mathrm{RT}}$ includes nation, birthday, gender, ZIP, the sensitive attribute is disease, $k = 2$. As can be seen from Table 1, $t1[\mathrm{QI}] = t2[\mathrm{QI}]$, $t3[\mathrm{QI}] = t4[\mathrm{QI}]$, $t5[\mathrm{QI}] = t6[\mathrm{QI}] = t7[\mathrm{QI}]$.

Definition 2. k-Degree anonymity. A social network graph $G(V\,; E)$ is said to satisfy k-degree anonymity, if each vertex (node) has $k - 1$ other vertexes at least, and these vertex's degree are same in the social network graph. The variable V represents vertex amounts, and E represents edge amounts between vertexes [16, 17].

Table 1: Example of 2-anonymity, $\mathrm{QI} = \{\mathrm{Nation, Birthday, Gender, ZIP}\}$.

No.	Nation	Birthday	Gender	ZIP	Salary
1	Yellow	1995	F	26118*	9000
2	Yellow	1995	F	26118*	5000
3	Yellow	1994	M	26112*	27000
4	Yellow	1994	M	26112*	10000
5	Brown	1993	F	26113*	7500
6	Brown	1993	F	26113*	30000
7	Brown	1993	F	26113*	9500

K-degree anonymity can prevent the inference attack by the adversary with background knowledge about vertex degree. In Figure 1, degree collection is $d = \{4, 3, 2, 4, 3, 3, 2, 3, 2\}$ in primal social network graph (a), so anonymity social network graph (b) satisfies 2-degree anonymity in Figure 1.

Definition 3. Graph isomorphism. For graphs: $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ where $\mid V_1 \mid = \mid V_2 \mid$, if there is a bijection h between V_1 and V_2 satisfies $\forall (u, v) \in E_1$, if and only if $\exists (h(u), h(v)) \in E_2$, $G1$, and G_2 are graph isomorphism, represented as $G_1 \cong G_2$. V_i represents vertex (node) numbers, and E_i represents edge numbers between vertexes.

For example, when we delete the node information of (a) and (b) in Figure 1, (a) and (b) are isomorphic [18].

Definition 4. k-Isomorphism. For a graph $G = (V, E)$, whose k sub-graphs are $g_1, g_2, \cdots, g_k$, if $g_i (l \le i \le k)$ satisfies: (1) $U_{i=1}^{k} g_i = G$; (2) $g_i \cap g_j = \Phi, (i \ne j)$ (3) g_i and $g_j (i \ne j)$ are isomorphism, then, the graph G is k-isomorphism.

Definition 5. k-Isomorphism vertex group. Given a k-isomorphism publishing graph $G_p = (\mathrm{VP, EP}) = \{g_1, g_2, \cdots g_k\}$, $\forall v_1 \in G_p, v_1 \in g_1$, then, there exist $k - 1$ vertexes $v_i \in g_i (i = 2, \cdots, k)$ are isomorphic to v_1, the vertex set consists the vertex v_i and the $k - 1 v_i \in (i = 2, \cdots, k)$ is k-isomorphism vertexes group, which is denoted as VCS, $|\mathrm{VCS}| = k$. Each VCS includes k vertexes and there are $|\mathrm{VP}|/k\,\mathrm{VCS}$ in the k-isomorphism graph G_p.

Definition 6. k-Isomorphism edge group. Given a k-isomorphism publishing graph $G_p = (\mathrm{VP, EP}) = \{g_1, g_2, \cdots, g_k\}$, $\forall$ $e_1 \in g_1$, then there exist $k - 1$ edges $e_i \in g_i (i = 2, \cdots, k)$ is isomorphic to e_1, the vertex set consists the vertex $e_{1,}$ and the $k - 1 e_i \in g_i (i = 2, \cdots, k)$ is k-isomorphism edges group, which is denoted as ECS, $|\mathrm{ECS}| = k$. Each VCS includes k edges and there are $|\mathrm{EP}|/kECS$ in the k-isomorphism graph G_p.

Definition 7. Social network graph. Given a social network graph: $G = (V, E)$, wherein vertex set V denotes the social individuals, and the edge set E denotes the relationships among the social individuals. Each vertex and edge has its identify and attribute which includes

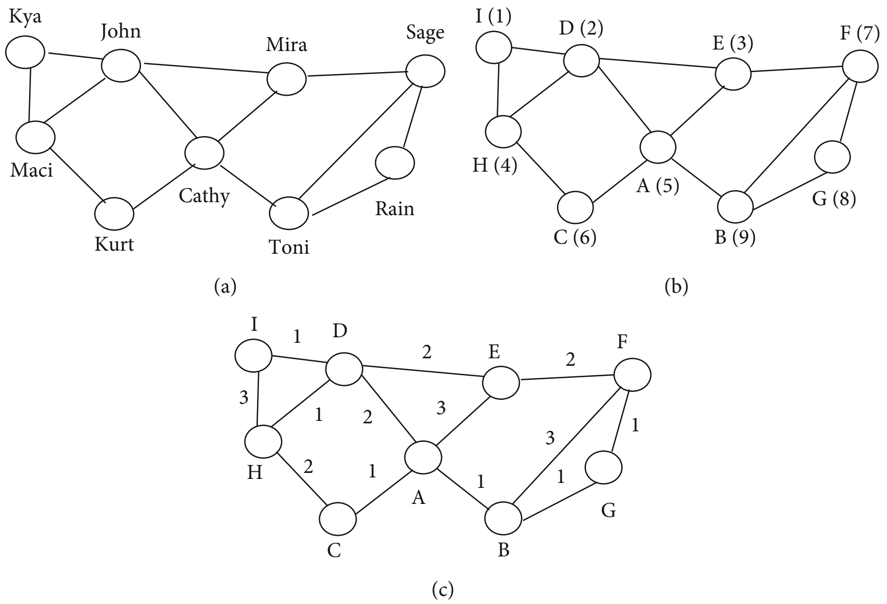


Figure 1: 2-Anonymity publishing in a social network.

(a) Identifier attribute (ID) of vertex as $v_i\{v_i^{\mathrm{ID}}\}$

(b) Quasi-identifier attribute (QI) of vertex v_i as $\mathrm{QI} = \{v_i^{N(1)}, \cdots, v_i^{N(s)}, v_i^{C(1)}, \cdots, v_i^{C(t)}\}$

(c) Sensitive attribute (SA) of vertex v_i as $\mathrm{SA} = \{v_i^{S(1)}\}$

(d) Quasi-identifier attribute (QI) of edge e_i as $\mathrm{QI} = \{e_j^{N(1)}, \cdots, e_j^{N(p)}, e_j^{C(1)}, \cdots, e_j^{C(q)}\}$

(e) Sensitive attribute (SA) of edge e_i as $\mathrm{SA} = \{e_j^{S(1)}\}$

(f) Other attributes (OA)

Attribute (QI) of edge denotes by vector pair $(v_i^{\mathrm{ID}}, v_j^{\mathrm{ID}})$, the total number of vertexes $N = |V|$, N denotes QI of numeric attribute, C denotes QI of character attribute, s, t, p, and q denote the amount of QI, respectively. For example, Figure 1 is an example of friendship social network, each vertex is a customer, and each edge denotes relationship between two vertexes. Table 2 is primal data of each vertex in Figure 1(a). Table 3 is edge table, Eid denotes the sequence number of edge, Vid1 and Vid2 denote the sequence number of vertex of Figure 1(b), and weighted relationship denotes the relationship between Vid1 and Vid2 of Figure 1(c). Table 4 is another relational data table of a vertex of Figure 1(a).

2.2. Sensitive Degree of Friend Relationship (SA) of Vertex and Edge

2.2.1. Sensitive Degree of Friend Relationship (SA) of Vertex. We use the influence matrix to represent the level of influence of vertex-sensitive attributes [19, 20]. We can use the

Table 2: The vertex table.

Vid	Name	Age	Gender	ZIP	Salary
1	Kya	25	F	261186	9000
2	John	25	M	261185	5000
3	Mira	26	F	261101	27000
4	Maci	26	F	261131	10000
5	Cathy	27	F	261131	7500
6	Kurt	27	M	261124	30000
7	Sage	27	F	261186	9500
8	Rain	37	M	261185	28000
9	Toni	35	M	261124	22000

Table 3: The edge table.

Eid	Vid1	Vid2	Weighted relationship
1	1	2	1
2	1	4	3
3	2	3	2
4	2	4	1
5	2	5	2
6	3	5	3
7	3	7	2
8	4	6	2
9	5	6	1
10	5	9	1
11	7	8	1
12	7	9	3
13	8	9	1

Table 4: Table of primal personal health information.

Name	Age	Gender	ZIP	Disease
Kya	25	F	261186	Cold
John	25	M	261185	Hypertension
Mira	26	F	261101	AIDS
Maci	26	F	261131	Cold
Cathy	27	F	261131	Cancer
Kurt	27	M	261124	Obesity
Sage	27	F	261186	Pneumonia
Rain	37	M	261185	Diabetes
Toni	35	M	261124	Short breath

influence matrix to meet the requirements of personalized privacy protection of users.

t_{ij}: the influence degree of NO. j sensitive attribute generated by NO. i vertex.

b_i: the weightiness of sensitive attribute value of NO. i vertex.

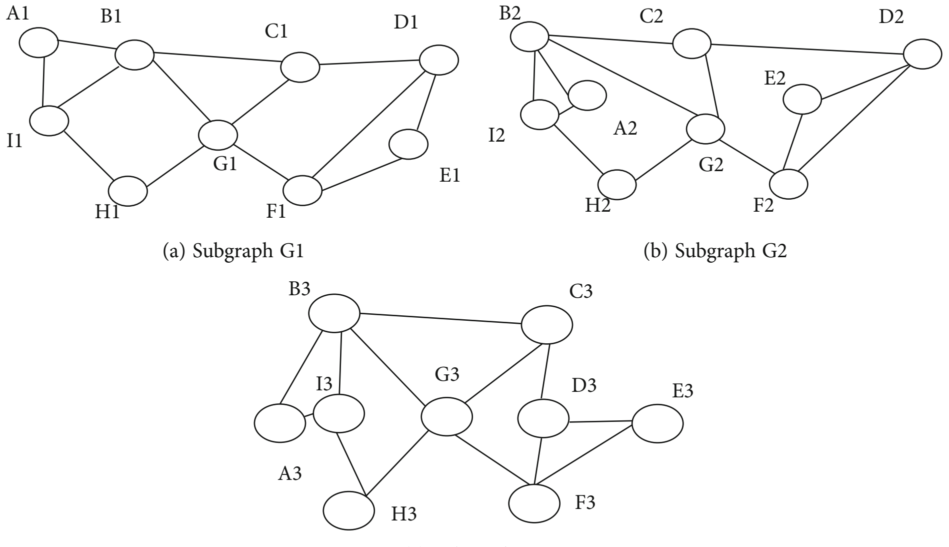
Influence matrix $|\mathrm{Vertex_m}|$ is with m rows, $n+1$ columns, m represents vertex amount, n represents QI attribute amount, so it can be described as

$$\mathrm{Vertex_m} = \left(t_{ij}|b_i\right)_{m\times(n+1)} \left[\begin{array}{ccccc|c} \mathrm{QI1} & \mathrm{QI2} & \mathrm{QI3} & \cdots & \mathrm{QI_n} & S \\ t11 & t12 & t13 & \cdots & \mathrm{t1_n} & b1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ t(m-1)1 & t(m-1)2 & t(m-1)3 & \cdots & t(m-1)_\mathrm{n} & bm-1 \\ tm1 & tm2 & tm3 & \cdots & tmn & bm \end{array}\right]. \quad (1)$$

The t_{ij}, b_i values come from experts or experience value. For example, the weightiness of QI in Table 4 can be divided into 5 grades, 1, 0.8, 0.4, 0.1, and 0, and the weightiness of S in Table 4 can be divided into 5 grades too, 0.10, 0.60, 0.70, 0.80, and 0.90. The cold is general disease, and disease weightiness value can use 0.1. Common cold (influenza) may have the character of a regional outbreak, and we define the weight value of the ZIP as 0.8. Common cold may also have a little bit to do with gender, and we define the weight value of gender as 0.1. Then, we define the disease weight values of obesity, short breath, hypertension, diabetes, pneumonia, cancer, and AIDS as 0.12, 0.31, 0.5, 0.6, 0.7, 0.91, and 0.92, respectively. The influence matrix is as follows according to Table 4.

$$\mathrm{Vertex_m} = \left(t_{ij}|b_i\right)_{9\times(5+1)} \left[\begin{array}{ccccc|c} \text{Nation} & \text{Occupation} & \text{Birthday} & \text{Gender} & \text{ZIP} & \text{Disease} \\ 0 & 0 & 0 & 0.1 & 0.8 & 0.1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0.1 & 0.4 & 0 & 0 & 0.6 \\ 0 & 0.1 & 0 & 0 & 0 & 0.31 \end{array}\right]. \quad (2)$$

2.2.2. Sensitive Degree of Friend Relationship (SA) of Edge. We described relationships of simple friend, good friend, and sweetheart friend (boyfriend or girlfriend) among the vertexes in Figure 1 of friend relationship graph. Graph (c) of Figure 1 is an example of friend relationship graph, "1" represents simple friend relationship between two vertexes, "2" represents good friend relationship between two vertexes, "3" represents sweetheart relationship between two vertexes, and "0" represents no relationship between two vertexes. Usually, if sweetheart friend includes gay or lesbian



(a) Subgraph G1

(b) Subgraph G2

(c) Subgraph G3

Figure 2: Isomorphism social relationships network graph.

Table 5: 9 vertex groups of 3-isomorphism.

VCS	$G1$	$G2$	$G3$
1	$A1$	$A2$	$A3$
2	$B1$	$B2$	$B3$
3	$C1$	$C2$	$C3$
4	$D1$	$D2$	$D3$
5	$E1$	$E2$	$E3$
6	$F1$	$F2$	$F3$
7	$G1$	$G2$	$G3$
8	$H1$	$H2$	$H3$
9	$I1$	$I2$	$I3$

relationship, most of people do not want others to know that he is gay or she is lesbian, so different people have different sensitive degree about friend relationships, so we must meet the needs of personalized privacy protection according to the practical application.

2.3. $\alpha(d,k)$-Anonymity Graph.

In order to make it impossible for an attacker to infer the real relationship between targeted individuals and corresponding vertexes with a probability, k-anonymity concept in data tables and the new concept of $\alpha(d,k)$-anonymity are introduced.

Definition 8. $\alpha(d,k)$-Anonymity of the vertex. Undirected graph $G = (V,E)$, the graph $G_p = (V_p, E_p)$ is as its anonymous publishing graph, if a vertex $v \in V$, there are at least $k-1$ vertexes $u_1, u_2, \cdots, u_{k-1} \in V_p$ in G_p, which makes $\text{Neighbor}_d(v) \cong \text{Neighbor}_d(u_i)$ and $v \neq u_i$, wherein, $i = 1, 2, \cdots k-1$, thus, the vertex v is (d,k)-anonymity, and the vertex v is $\alpha(d,k)$-anonymity according to α, α is the weight of relationships (edge weight) of d- neighborhood of vertex v.

Table 6: 13 edges groups of 3-isomorphism.

ECS	$G1$	$G2$	$G3$
1	$(A1, B1)$	$(A2, B2)$	$(A3, B3)$
2	$(A1, I1)$	$(A2, I2)$	$(A3, I3)$
3	$(B1, C1)$	$(B2, C2)$	$(B3, C3)$
4	$(B1, G1)$	$(B2, G2)$	$(B3, G3)$
5	$(B1, I1)$	$(B2, I2)$	$(B3, I3)$
6	$(C1, D1)$	$(C2, D2)$	$(C3, D3)$
7	$(C1, G1)$	$(C2, G2)$	$(C3, G3)$
8	$(D1, E1)$	$(D2, E2)$	$(D3, E3)$
9	$(D1, F1)$	$(D2, F2)$	$(D3, F3)$
10	$(E1, F1)$	$(E2, F2)$	$(E3, F3)$
11	$(F1, G1)$	$(F2, G2)$	$(F3, G3)$
12	$(G1, H1)$	$(G2, H2)$	$(G3, H3)$
13	$(H1, I1)$	$(H2, I2)$	$(H3, I3)$

For example, in Figures 1, $\alpha = \{1, 2, 3\}$ of vertex F (sage), and $\alpha = \{1, 2, 3\}$ of vertex H (Maci), so vertex F and vertex H satisfy $\alpha(1,2)$-anonymity.

Definition 9. $\alpha(d,k)$-Anonymity of the graph. Undirected graph $G = (V,E)$, the graph $G_p = (V_p, E_p)$ is as its anonymous publishing graph. If any vertex $v \in V$ is (d,k)-anonymity, thus, the graph G_p is (d,k)-anonymity, if any vertex $v \in V$ is $\alpha(d,k)$-anonymity, thus, the graph G_p is $\alpha(d,k)$-anonymity.

Definition 10. Individual information leakage. Suppose graph G_p is the anonymity publishing graph of social network graph G, when the relative sensitive coefficient k and l satisfy one of the following four conditions, then, there

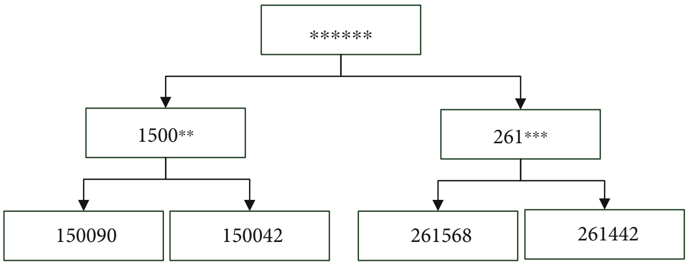


Figure 3: Hierarchies tree of ZIP.

exists individual information leakage. Otherwise, if the graph G_p can ensure that any of the following circumstances are not going to happen, the anonymity publishing graph G_p is regarded as secure. If the graph G_p can ensure the following circumstance (1) and (2) will not happen, then, the anonymity publishing graph G_p is k-secure [21].

(1) *Vertex Leakage*. The probability of ascertaining the corresponding relationship between the vertex in the graph G_p and the target individual A in the primal graph G is greater than $1/k$

(2) *Edge Leakage*. The probability of ascertaining the corresponding relationship between the edge in the graph G_p and the edge in the primal graph G is greater than $1/k$

(3) *Leakage of Vertex Sensitive Information*. The probability of ascertaining the sensitive information of target individual A in the primal graph G is greater than $1/l$

(4) *Leakage of Edge Sensitive Information*. The probability of ascertaining the sensitive information of the edge in the primal graph G is greater than $1/l$

2.4. Personalized (α, β, l, k)-Anonymity

2.4.1. Personalized (α, β, l, k)-Anonymity Model. Personalized (α, β, l, k)-anonymity satisfies the following conditions:

(1) Personalized (α, β, l, k)-anonymity satisfies $\alpha(d, k)$ -anonymity

(2) $\forall bi < \beta$ in matrix $|\text{Vertex}_\text{m}|$, all vertexes in k-isomorphism vertexes group be supposed to be published directly. Otherwise should be satisfied condition (3) and condition (4)

(3) $L = \sum_{j=1,\cdots,|k-I|} \text{count}(|bi - bj| > 0), 1 \le i \le |\text{VCS}|, bi, b$ j are S column vectors of influence matrix $|\text{Vertex}_\text{m}|$, $i \neq j$, L is the numbers of different sensitive attribute value

(4) If$P = \text{count}(|\text{MAX}_{i=1\cdots}|e_m|t_{ik}| = 1) > 0$ in influence matrix $|\text{Vertex}_\text{m}|$, when t_{ik} is generated, under the precondition of anonymity, promote generalization hierarchies, or suppress directly [19]. P denotes sensitive degrees between QI_k and S in influence matrix$|Vertexm|$, if $P = 1$, it means that t_{ik} will influence b_i's sensibility

Here, threshold β is important degree parameter of sensitive attribute in condition (2). If sensitive attribute values of an equivalent class (VCS) are less then β, that is to say sensitive attribute of these vertexes in k-isomorphism vertex group cannot affect their privacy, all vertexes can be published directly. Otherwise, must satisfy condition (3) and condition (4). If $L > 0$, number of different sensitive attribute value is greater than or equal to 2, L makes sensitive attribute diversity.

2.4.2. Personalized (α, β, l, k)-Anonymity Example. There is an example which is shown to explain the definition and the process of personalized (α, β, l, k)-anonymity according to Figure 2.

Figure 1(a) is the subgraph G of social relationships network, and the isomorphism subgraphs of G are found. The 3-isomorphism subgraphs are shown in Figure 2.

In Figure 2, (a) is the initial subgraph in Figure 1, and (b) and (c) are the isomorphism graphs corresponding to (a). From graph G, the amount of vertexes $|Vp|$ is 27, and the amount of edges $|Ep|$ is 39. Therefore, 9 3-isomorphism vertex groups and 13 3-isomorphosm edge groups are created and listed in Tables 5 and 6.

Now, the 9 3-isomorphism vertex groups are generalized by their identifier attributes according to parameter β. The isomorphism vertex groups VCS are changed into equivalence class vertexes groups QI. The item age, gender, and ZIP are identifier attributes, and disease item is the sensitive attribute. The inheritance hierarchy tree of ZIP is shown in Figure 3. The inheritance hierarchy tree of disease is shown in Figure 4 [21].

The $A1$, $A2$, and $A3$ attributes in the isomorphism groups VCS1 and VCS2 are listed in Table 7. After generalization, the identifier attributes value gen (VCS) are created and shown in Table 8 [15].

3. Personalized (α, β, l, k)-Anonymity Algorithm

The basic algorithm principle is that k-isomorphism graph $G_p = \{g_1, \cdots, g_e\}$ is caught; k-isomorphism graph vertex group VCS is generalized about identifier attributes and sensitive attributes; edge group ECS is generalized about identifier attributes and sensitive attributes. In the process, the

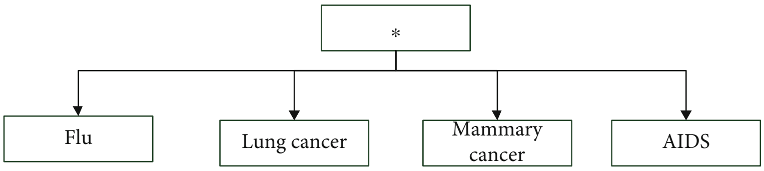


Figure 4: Hierarchies tree of disease.

Table 7: Example of isomorphism groups vertex's attributes values.

VCS	Num	Race	Occupation	Age	Gender	ZIP	Disease
1	$A1$	Asian	Salesman	25	M	150086	Flu
	$A2$	Asian	Salesman	35	F	150084	Flu
	$A3$	Black	Teacher	35	M	150081	Mammary cancer
2	$B1$	White	Teacher	45	F	150090	Lung cancer
	$B2$	White	Driver	45	M	150041	Lung cancer
	$B3$	White	Driver	50	M	150024	Lung cancer

Table 8: Example of isomorphism groups' generalization identifier attributes values.

VCS	Num	Race	Occupation	Age	Gender	ZIP	Disease
1	$A1$	Asian	Salesman	[25,35)	*	15008*	Flu
	$A2$	Asian	Salesman	[25,35)	*		Flu
	$A3$	Black	Teacher	[25,35)	*		Mammary cancer
2	$B1$	White	Teacher	[45,50)	F	1500**	*
	$B2$	White	Driver	[45,50)	M		*
	$B3$	White	Driver	[45,50)	M		*

generalization is not executed definitely, especially when the type differences do not affect l-diversity [22]. The input parameter $\alpha = \{0, 1, 2, 3\}$ indicates the generalizing type: when the value is 0, it should be static generalization, and when the value is greater than 0, it means the generalizing would be on the base of graph isomorphism. The input parameter α indicates the sensitive degree between nodes. When $\alpha \neq 0$, we achieve $\alpha(d, k)$-anonymity graph, d-neighborhood attack of graph and structure attack of graph can be prevented [23, 24], when $\alpha = 0$, the input parameter β is the generalization threshold [19, 22], background knowledge attack and homogeneity attack can be prevented by using anonymous data of vertexes in social network effectively, and diversity of sensitive attribute can be solved. The following is personalized (α, β, l, k)-anonymity algorithm $(\alpha = 0)$, and personalized (α, β, l, k)-anonymity algorithm $(\alpha \neq 0)$ has been given in another paper published by the author [23].

4. Experiments and Results

The experiments were completed in the PC with Intel(R) Core(TM) i5-4590 CPU @ 3.30 GHz, 8 GB memory, and the OS is Microsoft Windows 7. The programs were coded and compiled in VS 2019 IDE.

The vertex (nodes) data set in these experiments are from adults census data set of the UC Irvine Machine Learning Repository [25, 26]. There are two experiments examples, and the vertex numbers of each are 300 and 1000. In these vertexes, 6 attributes are considered in the experiments, which are age, occupation, race, gender, zip, and disease. In these attributes, age is numeric, and the others are category. The attribute disease is sensitive attribute. The edge set in these experiments is created by Pajek software randomly, and the numbers of nodes are, respectively, 5000, 10000, 15000, 20000, and 25000.

Information loss was compared between the algorithm in this paper that we proposed and paper [15]. We use the information loss method from paper [15]. The algorithm in this paper was named as ACIM (anonymous composite improved model) algorithm, and the algorithm in paper [15] was named as ACM (anonymous composite model) algorithm. In personalized (α, β, l, k)-anonymity algorithm $(\alpha = 0)$, we make the data usability and original according to parameter β. When β is less than the given threshold, all vertex (data) will be published directly, which reduce the degree of data distortion [19].

```
Inputs:
Initial anonymous graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$,
Sensitivity parameters: $\boldsymbol{k'}(\boldsymbol{k'} \geq 2)$; $\boldsymbol{l}(2 \leq \boldsymbol{l} \leq \boldsymbol{k'})$; $\boldsymbol{m}(\boldsymbol{l} \leq \boldsymbol{m} \leq |\boldsymbol{V}|)$;
Node attributes table: $\boldsymbol{AS} = \{\boldsymbol{v_i^S}, \boldsymbol{v_i^{N(1)}}, \ldots, \boldsymbol{v_i^{N(s)}}, \boldsymbol{v_i^{C(1)}}, \ldots, \boldsymbol{v_i^{C(t)}}\}$;
Edge attributes table: $\boldsymbol{AS} = \{\boldsymbol{v_j^S}, \boldsymbol{v_j^{N(1)}}, \ldots, \boldsymbol{v_j^{N(s)}}, \boldsymbol{v_j^{C(1)}}, \ldots, \boldsymbol{v_j^{C(t)}}\}$;
All the classified attribute inheritance tree $\boldsymbol{H_C}$
Input parameters $\boldsymbol{\alpha = 0}$ and $\boldsymbol{\beta}$
Outputs:
Anonymous graph $\boldsymbol{Gp} = \{\boldsymbol{g_1}, \boldsymbol{g_2}, \ldots, \boldsymbol{g_e}\}$;
The whole VCS, ECS and their attribute information;
Steps:
anonymous graph $\boldsymbol{Gp}$, groups VCS and ECS are caught;
read $\alpha$ to judge the generalizing type;
got the group number:$\boldsymbol{N_{VCS}} = |\boldsymbol{N_{VP}}|/\boldsymbol{k}$, $\boldsymbol{N_{ECS}} = |\boldsymbol{N_{EP}}|/\boldsymbol{k}$;
for$i = 1$ to $N_{VCS}$do//$QI$ attributes generalization
  for$j = 1$ to $s$do//numeric type attributes generalization
    $\boldsymbol{gen(VCS_i)[N_j]} = [min\{v_1^{N(j)}, \ldots, v_k^{N(j)}\}, max\{v_1^{N(j)}, \ldots, v_k^{N(j)}\}]$
    end for
    for$j = 1$ to $t$do//$t$ type $QI$ attributes generalization
      $gen(VCS_i)[C_j] = \{v_1^{C(j)}, \ldots, v_k^{C(j)}\}$
    end for
    while$(|SA(VCS_i)| < l \&\& |SA(VCS_i)|/|SA(VP)| > \boldsymbol{\beta})$do
      if(sensitive attributes are classified) then
        for$j = 0$ to $k$do
          $v_j^C$ is replaced by its parents node in classified inheritance tree of sensitive attribute
          if$(|SA(VCS_i) \geq l|)$then jump while loop
        end for
      Else
        for$j = 1$ to $k$do
          the interval of $v_j^N$ is changed to its neighborhood;
          if$(|SA(VCS_i) \geq l|)$then jump while loop
        end for
      end if
    end while
  end for
  for$i = 1$ to $N_{ECS}$do
    for$j = 1$ to $p$do
    $\boldsymbol{gen(ECS_i)[N_j]} = [min\{v_1^{N(j)}, \ldots, v_k^{N(j)}\}, max\{v_1^{N(j)}, \ldots, v_k^{N(j)}\}]$
    end for
    for$j = 1$ to $q$do
      $gen(ECS_i)[C_j] = \{e_1^{C(j)}, \ldots, e_k^{C(j)}\}$
    end for
    while$(|SA(VCS_i)| < l \&\& |SA(VCS_i)|/|SA(V_P)| > \boldsymbol{\beta})$do
      if(sensitive attributes are classified) then
        for$j = 1$ to $k$do
          $v_j^C$ is replaced by its parents node in classified inheritance tree of sensitive attribute
          if$(|SA(VCS_i) \geq l|)$then jump while loop
        end for
      Else
        for$j = 1$ to $k$do
          the interval of $n_j^N$ is changed to its neighborhood;
          if$(|SA(VCS_i) \geq l|)$then jump while loop
        end for
      end if
    end while
end for
anonymous graph $\boldsymbol{G_p}$ is published; all the $VCS$ nodes, $ECS$ edges and their attribute information are published
```

Algorithm 1: Personalized (α, β, l, k)-anonymity algorithm.

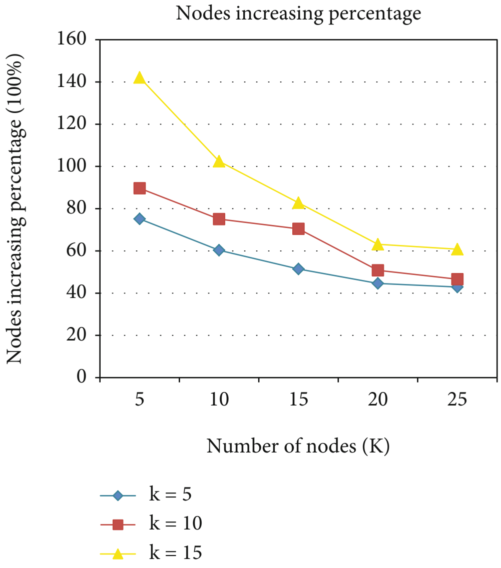


FIGURE 5: Nodes increasing percentage.

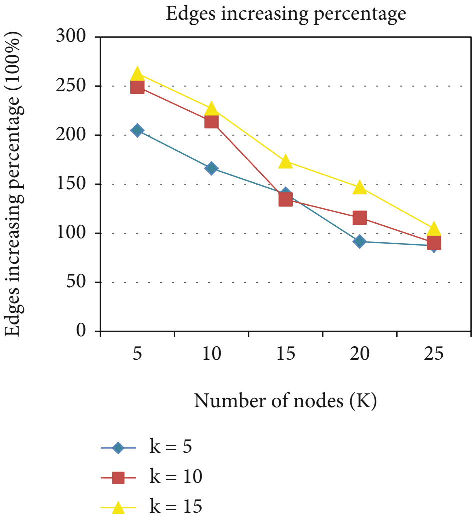


FIGURE 6: Edges increasing percentage.

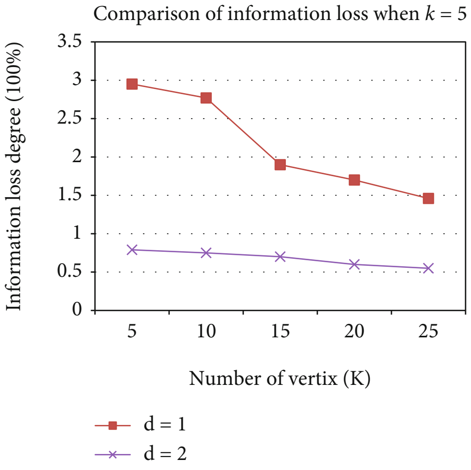


FIGURE 7: Information loss degree when $k = 5$, $d = 1$, and $d = 2$.

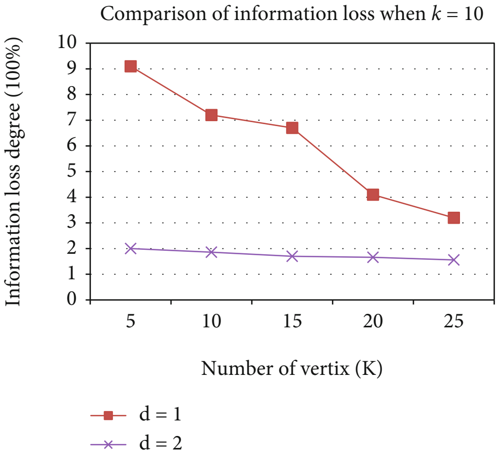


FIGURE 8: Information loss degree when $k = 10$, $d = 1$, and $d = 2$.

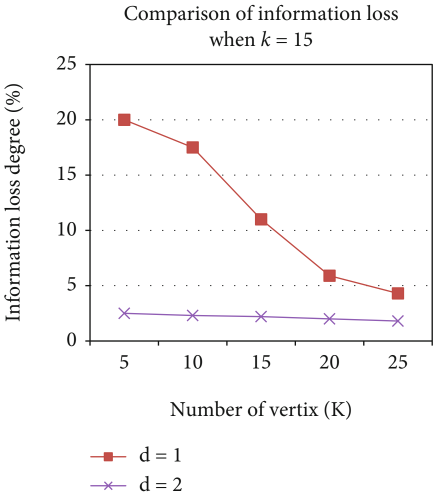


FIGURE 9: Information loss degree when $k = 15$, $d = 1$, and $d = 2$.

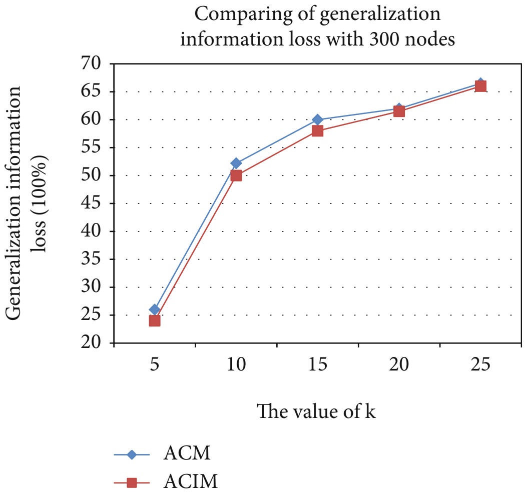


FIGURE 10: Comparison of generalization information loss with 300 nodes.

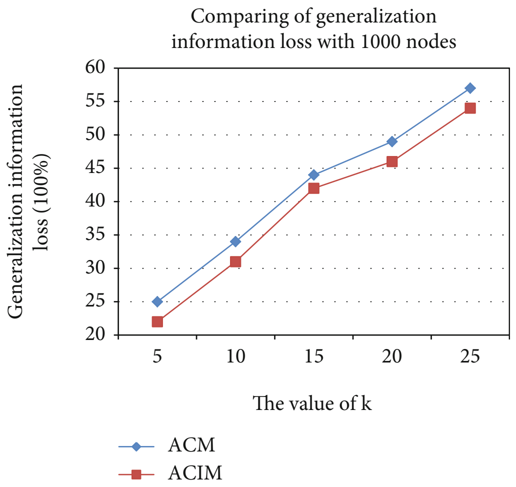


Figure 11: Comparison of generalization information loss with 1000 nodes.

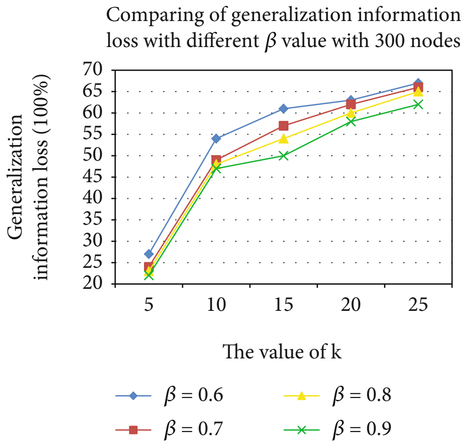


Figure 12: Comparison of generalization information loss with different β value with 300 nodes.

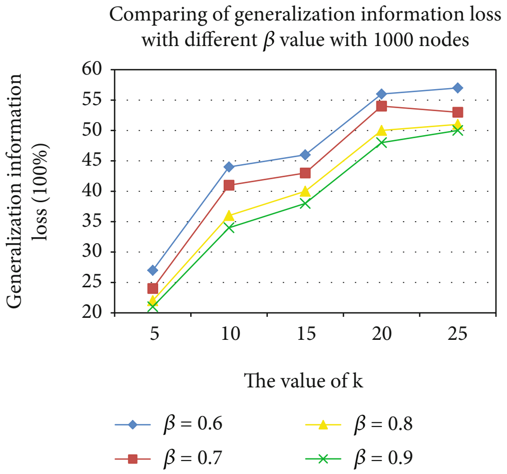


Figure 13: Comparison of generalization information loss with different β value with 1000 nodes.

When $\alpha \neq 0$, a number of nodes and edges should be added in initial graphs. When the structure is more different, the number of adding is higher. Meanwhile, the information loss is larger.

Figure 5 shows that some nodes are added to construct the isomorphic graphs, the percentage of adding nodes in all the nodes of the graph is shown in Figure 5, and the situation of edges is shown in Figure 6. With the $k = 5$, $k = 10$, and $k = 15$, the increasing speed of nodes and edges slows down. These are additional redundant data.

In Figures 7–9, the loss of information is shown with $k = 5$, $k = 10$, and $k = 15$. The information loss degrees are increasing with the increasing of nodes, k and d. The reason is that the candidate set will be larger with the increasing of data scale, and finding similar neighborhood will be easier [19].

When $\alpha = 0$, the information loss results of attributes generalization are compared when k value is 5, 10, 15, 20, and 25. Figures 10 and 11 show the comparison results.

From Figure 10, when k value increase, the demand of privacy protection becomes higher, which lead to obviously increasing of information loss. Besides, the loss of ACIM is lower after comparison. The reason is that not all the situations in same types are generalized but adding threshold judgments in ACIM. In Figure 11, the number of nodes is larger, and the generalization information loss is lower and make the node information availability for users.

Figures 12 and 13 show the comparison of generalization information loss with different β value. With higher β value, the vertex's attributes should be changed less, so the information loss rate should be lower. That is to say, parameter β can make the vertex information availability and meet personalized needs.

5. Conclusion

The authors study k-anonymity technologies and introduce k-anonymity application in relational database and social network. We proposed personalized (α, β, l, k)-anonymity model of social network. A lot of personalized (α, β, l, k)-anonymity algorithm experiments were done by the authors. Experimental results show that d-neighborhood attack of graph, background knowledge attack, and homogeneity attack can be prevented effectively by using anonymous vertexes and edges, as well as the influence matrix based on background knowledge. The diversity of vertex-sensitive attribute can be achieved. Personalized protecting privacy requirements can be met by using such parameter as α, β, l, k.

Data Availability

Previously reported data were used to support this study and are available at Adult Data Set of the UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets/Adult.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors want to thank the helpful comments and suggestions from the anonymous reviewers. This work was supported in part by the Natural Science Foundation of Taizhou University (Grant no. TZXY2019QDJJ008) and the Natural Science Foundation of Heilongjiang Province of China (Grant no. JJ2019LH0048).

References

[1] X. M. Ren, B. X. Jia, K. C. Wang, and J. Cheng, "Research on k-anonymity privacy protection of social network," *Applied Mechanics and Materials*, vol. 530-531, pp. 701–704, 2014.

[2] H. Miyajima, N. Shigei, H. Miyajima, Y. Miyanishi, S. Kitagami, and N. Shiratori, "New privacy preserving clustering methods for secure multiparty computation," *International Journal of Computer Science*, vol. 6, no. 1, pp. 270–276, 2016.

[3] K. Macwan and S. Patel, "Privacy preservation approaches for social network data publishing," in *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*, pp. 213–233, Springer, Cham, 2021.

[4] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge Discovery and Data Mining(KDD)*, pp. 611–617, Philadelphia, Pennsylvania, USA, 2006.

[5] Y. J. Luo, Q. Liu, and Y. Wang, "Overview of protecting user privacy in social networks," *Application Research of Computers*, vol. 10, pp. 3061–3064, 2010.

[6] Y. Xiaowei and Z. Weitao, "On link privacy in randomizing social networks," *Knowledge and Information Systems*, vol. 28, no. 3, pp. 645–663, 2011.

[7] J. Cheng, A. W. Fu, and J. Liu, "K-isomorphism: privacy preserving network publication against structural attacks," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pp. 459–470, Indianapolis, Indiana, USA, 2010.

[8] K. Liu and E. Terzi, "Towards identity anonymization on graphs//proceedings of the 2008 ACM SIGMOD international conference on management of data," in *ACM*, pp. 93–106, Association for Computing Machinery, New York, NY, 2008.

[9] A. Campan and T. Truta, "Data and structural k-anonymity in social networks," *Privacy, Security, and Trust in KDD*, vol. 5456, pp. 33–54, 2009.

[10] M. K. SUNG, K. Y. LEE, J.-B. SHIN, and Y. D. CHUNG, "A privacy protection method for social network data against content/degree attacks," *IEICE Transactions on Information and Systems*, vol. E95-D, no. 1, pp. 152–160, 2012.

[11] E. Y. Baagyere, Z. Qin, H. Xiong, and Q. Zhiguang, "The structural properties of online social networks and their application areas," *International Journal of Computer Science*, vol. 43, no. 2, pp. 270–276, 2016.

[12] N. Li and X.-L. Zhang, "Research on dynamic social network anonymity technology for protecting community structure," *International Journal of Network Security*, vol. 23, no. 4, pp. 576–587, 2021.

[13] X. K. Xiao and Y. F. Tao, "Personalized privacy preservation," in *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pp. 229–240, Chicago, Illinois, USA, 2006.

[14] X. Zhang, J. Liu, H. Bi, J. Li, and Y. Wang, "Personalized K-in&out-degree anonymity method for large-scale social networks based on hierarchical community structure," *International Journal of Network Security*, vol. 23, no. 2, pp. 314–325, 2021.

[15] L. Sweeney, "k-Anonymity: a model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[16] H. Wu, J. Zhang, B. Wang, J. Yang, and B. Sun, "(d, k)-Anonymity for social networks publication against neighborhood attacks," *Journal of Convergence Information Technology*, vol. 8, no. 2, pp. 59–67, 2013.

[17] H. W. Wu, *Research on Anonymity Techniques for Privacy-Preserving Data Publishing in Social Networks*, Harbin Engineering University dissertation for the Degree of Doctor, 2013.

[18] L. Zou, L. Chen, and M. T. Ozsu, "k-Automorphism," *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 946–957, 2009.

[19] X. Ren, J. Yang, and F. Wei, "Research on CBK(L,K)-anonymity algorithm," *International Journal of Advancements in Computing Technology*, vol. 3, no. 4, pp. 165–173, 2011.

[20] L. I. Siyu, *Research for Protecting Privacy of Social Network Data Based on Relevance Degree Perception*, Guangxi Normal University, 2020.

[21] P. Samarati and L. Sweeney, "Generalizing data to provide anonymity when disclosing information," *Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, p. 188, 1998.

[22] B. Zhou and J. Pei, "The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks," *KAIS*, vol. 28, no. 1, pp. 47–77, 2011.

[23] X. M. Ren, D. X. Jiang, K. C. Wang, and R. A. N. Qi, "A personalized a (d, k)-anonymity for social network," in *2017 2nd international conference on computer, Mechatronics and Electronic Engineering*, pp. 167–174, Singapore, 2017.

[24] B. Zhou and J. Pei, "Preserving privacy in social networks against neighborhood attacks," in *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, ICDE'08*, pp. 506–515, Cancun, Mexico, 2008.

[25] D. J. Newman, S. Hettich, C. L. Blake, and C. J. Merz, *UCI Repository of Machine Learning Databases*, 1998, http://archive.ics.uci.edu/ml/datasets/Adult.

[26] J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A. W. C. Fu, "Utility-based anonymization for privacy preservation with less information loss," *Journal of SIGKDD Explorations*, vol. 8, no. 2, pp. 21–30, 2006.