Research on \((\alpha, \beta, l, k)\)-anonymity Model for Social Network

Xiang-min REN, De-xun JIANG, Ke-chao WANG and Qi RAN

School of Software, Harbin University, No.109, Zhongxing Road, Harbin, China

Keywords: Social network, Privacy preserving, \((\alpha, \beta, l, k)\)-anonymity model.

Abstract. More and more people pay attention to the security problems in social network. We propose a \((\alpha, \beta, l, k)\)-anonymity model to protect the publishing data security based on the graph structure of social network. It can make anonymous vertexes and edges of graph effectively resist \(d\)-neighborhood attack, and structure attack. At the same time, it can make anonymous data of vertexes in social network effectively resist background knowledge attack and homogeneity attack by using the influence matrix of background knowledge, and can solve diversity of vertexes sensitive attribute, meet personalized needs according to the weighted parameter \(\alpha, \beta\) given in advance.

Introduction

More and more people make friends, share information, watch video, play games on social network such as Facebook, Twitter, Wechat and so on. The sensitive or private information(data) of social network are possibly leaked at any moments when publishing and sharing the information(data). The data in social network is more complex than two dimensional data table(relational database)[1], we cannot use anonymity methods and technologies, which can be used into relational database, into social network directly. For example, Table 1 adheres to \(k\)-anonymity based on relational database, the quasi-identifier includes Race, Birth, Sex, ZIP, the sensitive attribute is Disease, \(k=2\). In particular, \(t_1[QI]= t_2[QI], t_3[QI]= t_4[QI], t_5[QI]= t_6[QI]= t_7[QI]\). So, scholars begin to research and develop the publishing data security technologies of social network.

| Num | Race | Birth | Sex | ZIP    | Disease    |
|-----|------|-------|-----|--------|------------|
| t1  | white| 1983  | M   | 8460** | flu        |
| t2  | white| 1983  | M   | 8460** | short breath |
| t3  | white| 1983  | F   | 8459** | diabetes   |
| t4  | white| 1983  | F   | 8459** | chest pain |
| t5  | brown| 1984  | M   | 8459** | pneumonia  |
| t6  | brown| 1984  | M   | 8459** | obesity    |
| t7  | brown| 1984  | M   | 8459** | hypertension |

Problem Definition

Related Concepts

Definition 1. \(k\)-degree Anonymity. A social network graph \(G(V;E)\) is \(k\)-degree anonymity, if each vertex(node) at least has \(k-1\) other vertexes(nodes) with the same degree in the same social network graph. \(V\) is numbers of vertexes(nodes), \(E\) is numbers of edges between vertexes [2][3].

\(K\)-degree anonymity can resist the adversary who has background knowledge of vertex degree. In Figure 1, the collection of degree is \(d=\{4,3,2,4,3,2,3,2\}\) in original social network graph (a), so social network graph (b) satisfies 2-degree anonymity in Figure 1.
Definition 2. Graph Isomorphism. For two graphs \(G_1 = (V_1, E_1)\) and \(G_2 = (V_2, E_2)\) where \(|V_1| = |V_2|\), if there is a bijection \(h\) between \(V_1\) and \(V_2\) satisfies: \(\forall (u, v) \in E_1\), if and only if \(\exists (h(u), h(v)) \in E_2\), \(G_1\) and \(G_2\) are graph Isomorphism, denoted as \(G_1 \cong G_2\). \(V_i\) is numbers of vertexes(nodes), \(E_i\) is numbers of edges between vertexes.

For example, when we delete the nodes information of (a) and (b) in Figure 1, (a) and (b) are isomorphic [4].

Definition 3. \(k\)-Isomorphism Vertex Group. Given a \(k\)-Isomorphism publishing graph \(G_p = (VP, EP) = \{g_1, g_2, \ldots, g_k\}\), \(\forall v_i \in G_p\), \(v_i \in g_i\), then there exist \(k-1\) vertexes \(v_i \in g_i\) \((i = 2, \ldots, k)\) is isomorphic to \(v_1\), the vertex set consists the vertex \(v_1\) and the \(k-1\) \(v_i \in g_i\) \((i = 2, \ldots, k)\) is \(k\)-Isomorphism vertexes group, which is denoted as \(VCS, |VCS| = k\). Each \(VCS\) includes \(k\) vertexes and there are \(|VP|/k\) \(VCS\) in the \(k\)-Isomorphism graph \(G_p\).

Definition 4. \(k\)-Isomorphism Edges Group. Given a \(k\)-Isomorphism publishing graph \(G_p = (VP, EP) = \{g_1, g_2, \ldots, g_k\}\), \(\forall e_i \in E_p\), \(\forall e_i \in g_i\), then there exist \(k-1\) edges \(e_i \in g_i\) \((i = 2, \ldots, k)\) is isomorphic to \(e_1\), the vertex set consists the vertex \(e_1\) and the \(k-1\) \(e_i \in g_i\) \((i = 2, \ldots, k)\) is \(k\)-Isomorphism edges group, which is denoted as \(ECS, |ECS| = k\). Each \(VCS\) includes \(k\) edges and there are \(|EP|/k\) \(ECS\) in the \(k\)-Isomorphism graph \(G_p\).

Definition 5. Social Network Graph. Given a graph \(G = (V, E)\), where vertex set \(V\) represents the social individual in the social network and the edge set \(E\) represents the relationship between the social individuals. Each vertex and edge has its identify and attribute which includes:

(a) Identifier attribute(\(ID\)) of vertex as \(v_i\) \(\{v_i^{ID}\}\);
(b) Quasi-identifier attribute(\(QI\)) of vertex \(v_i\) as \(QI = \{v_i^{N(i)}, v_i^{C(i)}\}\);
(c) Sensitive attribute(\(SA\)) of vertex \(v_i\) as \(SA = \{v_i^{S(i)}\}\);
(d) Quasi-identifier attribute(\(QI\)) of edge \(e_i\) as \(QI = \{e_i^{N(i)}, e_i^{C(i)}\}\);
(e) Sensitive attribute(\(SA\)) of edge \(e_i\) as \(SA = \{e_i^{S(i)}\}\);
(f) Other attributes(\(OA\)).

Table 2. The vertex table of Figure 1 (b).

| Vid | Name  | Age | Sex | ZIP    | Salary |
|-----|-------|-----|-----|--------|--------|
| 1   | Ken   | 34  | F   | 150086 | 8000   |
| 2   | Jone  | 38  | M   | 150085 | 4000   |
| 3   | Mars  | 25  | F   | 150001 | 17000  |
| 4   | Mary  | 43  | F   | 150031 | 9000   |
| 5   | Cindy | 27  | F   | 150031 | 6500   |
| 6   | Kite  | 19  | M   | 150024 | 20000  |
| 7   | Sarah | 25  | F   | 150086 | 8500   |
| 8   | Rachel| 37  | M   | 150085 | 18000  |
| 9   | Tom   | 20  | M   | 150024 | 12000  |
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 original personal health information of Figure 1(a).

| Name | Age | Sex | ZIP       | Disease    |
|------|-----|-----|-----------|------------|
| Ken  | 34  | F   | 150086    | hypertension |
| Jone | 38  | M   | 150085    | AIDS       |
| Mars | 25  | F   | 150001    | AIDS       |
| Mary | 43  | F   | 150031    | flu        |
| Cindy| 27  | F   | 150031    | cancer     |
| Kite | 19  | M   | 150024    | obesity    |
| Sarah| 25  | F   | 150086    | pneumonia  |
| Rachel| 37  | M   | 150085   | diabetes   |
| Tom  | 20  | M   | 150024    | short breath |

Attribute (QI) of edge denotes by vector pair \( (v_i^{ID}, v_j^{ID}) \), the total number of vertexes \( N = |\mathcal{V}| \), \( N \) denotes QI of numeric attribute, \( C \) denotes QI of character attribute, \( s, t, p, q \) denote the amount of QI respectively. For example, Figure 1 is an example of friendship social network, each vertex is a customer, each edge denotes relationship between two vertexes, Table 2 is original data of each vertex in Figure 1.

**Sensitive Degree of Friend Relationship (SA) of Vertex and Edge**

**Sensitive Degree of Friend Relationship (SA) of Vertex.** We use the influence degree of sensitive attribute of vertex produced by QI attributes and sensitive attribute itself according to [5]:

- \( t_{ij} \): the influence degree of NO. \( j \) sensitive attribute produced by NO. \( i \) QI attribute.
- \( b_i \): the weight of sensitive attribute value of NO. \( i \).

Influence matrix \( |\text{Vertex}| \) is with \( m \) rows and \( n+1 \) columns, \( m \) is the number of vertexes, \( n \) is the number of QI attribute, then the sample matrix is as follows:

\[
\text{Vertex} = (t_{ij} b_i)_{m \times (n+1)}
\]

For example, we can divide weight of QI into 5 grades: 1, 0.8, 0.6, 0.4, 0.2, and divide weight of S into 5 grades: 0.10, 0.20, 0.30, 0.40, 0.50 according to Table 4. The flu is common ailments, and disease weight can use 0.11, because of the characteristic of local outbreaks of flu, ZIP weight use 0.8, Sex weight use 0.2 etc. The disease weight of obesity, short breath, hypertension, diabetes, pneumonia, cancer and AIDS are 0.12, 0.21, 0.30, 0.31, 0.42, 0.51 and 0.52 respectively, different disease must have different disease weight value.
Sensitive Degree of Friend Relationship(SA) of Edge. We described relationships of simple friend, good friend, sweetheart friend (boyfriend or girlfriend) as “1”,”2”,”3” sensitive degree respectively among the vertexes in Figure 1 of friend relationship graph. Graph (c) of Figure 1 is an example of friend relationship graph, sweetheart friend is sensitive friend relationship for some people because they do not want others to know that he(she) has girlfriend(boyfriend).

\[ \alpha(d,k) \]-anonymity of the Vertex

Definition 6 \( \alpha(d,k) \)-anonymity of the Vertex. For an 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 \( \text{Neighbor}_v(u_i) = \text{Neighbor}_v(u_j) \) 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 \( \alpha \), \( \alpha \) is the weight of relationships(edge weight) of \( d \)-neighborhood of vertex \( v \).

For example, in Figure 1(c), \( \alpha=\{1,2,3\} \) of vertex F(Sarah), and \( \alpha=\{1,2,3\} \) of vertex H(Mary), so vertex A and vertex H satisfy \( \alpha(1,2) \)-anonymity.

Personalized \( (\alpha,\beta,l,k) \)-anonymity

Personalized \( (\alpha,\beta,l,k) \)-anonymity Model. Personalized \( (\alpha,\beta,l,k) \)-anonymity satisfies the following conditions:

1. Personalized \( (\alpha,\beta,l,k) \)-anonymity satisfies \( \alpha(d,k) \)-anonymity;
2. \( \forall b<\beta \) in matrix \( \left| \text{Vertex}_a \right| \), all vertexes in \( k \)-Isomorphism vertexes group should be published directly. Otherwise must satisfy condition (3) and condition (4);
3. \( L = \sum_{i=1}^{n} \text{count}(|b_i - b_j| > 0, 1 \leq i \leq |VCS|, b_i, b_j \text{ are column vector in influence matrix}| \text{Vertex}_a|, L \) is the amounts of different sensitive attribute value;
4. If \( P = \text{count}(| \text{MAX}_{t \in T} \text{count}(|t| = 1) > 0 \in \text{influence matrix}| \text{Vertex}_a|, \) when \( t_a \) is generated, under the precondition of anonymity, promote generalization hierarchies or suppress directly.

Here threshold \( \beta \) is important degree parameter of sensitive attribute in condition (2). If sensitive attribute values of a equivalent class (VCS) are less then \( \beta \), that is to say sensitive attribute of these vertexes in \( k \)-Isomorphism vertexes group can’t 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.

The Example of Personalized \( (\alpha,\beta,l,k) \)-anonymity. One example is shown to explain the definition and the process of personalized \( (\alpha,\beta,l,k) \)-anonymity. Figure 1 is the sub-graph \( G \) of social relationships network, and the isomorphism sub-graphs of \( G \) are found. The \( 3 \)-isomorphism sub-graphs are shown in Figure 2.

\[ \begin{align*}
\text{(a) Sub-graph G1} & \\
\text{(b) Sub-graph G2} & \\
\text{(c) Sub-graph G3} & 
\end{align*} \]

Figure 2. Isomorphism social relationships network graph.

In Figure 2, (a) is the initial sub-graph in Figure 1, and (b) and (c) are the isomorphism graphs corresponding to (a). From graph \( G \) the number of vertexes \( |V| \) is 27, the number of edges \( |E_p| \) is 39. Therefore, 9 \( 3 \)-isomorphism vertexes groups and 13 \( 3 \)-isomorphosm edges groups are created and listed in Table 5 and Table 6.
Table 5. Vertexes 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   |

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)  |

Now the 9 isomorphism vertexes groups are generalized by their identifier attributes according to parameter \(\beta\). The isomorphism vertexes groups VCS are changed into equivalence class vertexes groups \(QI\). The item Age, Sex, ZIP code are identifier attributes, and Disease item is the sensitive attribute. 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.

Table 7. Example of isomorphism groups vertex’s attributes values.

| VCS | Num | Race | Occupation | Age | Sex | ZIP     | Disease          |
|-----|-----|------|------------|-----|-----|---------|------------------|
| 1   |     | asian| Salesman   | 25  | M   | 150086  | flu              |
|     | A1  |      | Salesman   | 35  | F   | 150084  | flu              |
|     | A2  |      | Teacher    | 35  | M   | 150081  | mammary cancer   |

Table 8. Example of isomorphism groups’ generalization identifier attributes values.

| VCS | Num | Race | Occupation | Age | Sex | ZIP     | Disease          |
|-----|-----|------|------------|-----|-----|---------|------------------|
| 1   |     | asian| Salesman   | [25,35] | * | 15008* | flu              |
|     | A1  |      | Salesman   | [25,35] | * | 15008* | flu              |
|     | A2  |      | Teacher    | [25,35] | * | mammary cancer |

Conclusions

We proposed personalized \((\alpha,\beta, l, k)\)-anonymity mode. It can make anonymous vertexes effectively resist \(d\)-neighborhood attack of graph, and structure attack of graph. At the same time, it can make anonymous data of vertexes in social network effectively resist background knowledge attack and homogeneity attack, and can solve diversity of sensitive attribute, meet personalized needs according to the weighted parameter \(\alpha,\beta\) given in advance. We will give experiment results in another paper because the layout restrictions.
Acknowledgement

This work was supported in part by “Application Technology Research and Development of Harbin” (Grant No. 2014RFQXJ062), the project “Youth Academic Backbone of Ordinary Higher School of Heilongjiang Province” (Grant No. 1254G037), “Research Fund for the Innovative Scholars of Harbin”(Grant No.2016RAQXJ013)

References

[1] L. Sweeney, “k-anonymity: a model for protecting privacy”, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, no. 5, pp. 557-570, 2002.

[2] H.W. Wu, J.P. Zhang, B. Wang, et al., “(d, k)-anonymity for social networks publication against neighborhood attacks”, Journal of Convergence Information Technology, vol. 8(2), pp. 59-67, 2013.

[3] 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.

[4] L. Zou, L. Chen, M.T. Ozsu, “k-automorphism: General Framework for Privacy Preserving Network Publication”, Proceedings of the VLDB Endowment, vol. 2, pp. 946-957, 2009.

[5] X.M. Ren, J. Yang, et al., “Research on CBK(L,K)-Anonymity Algorithm”, International Journal of Advancements in Computing Technology, Vol. 3, No. 4, pp. 165-173, 2011.