

# A $k$ -Anonymization Method for Social Network Data with Link Prediction

Risa Sugai, Yuichi Sei<sup>a</sup>, Yasuyuki Tahara<sup>b</sup> and Akihiko Ohsuga<sup>c</sup>

*Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan*

**Keywords:** Social Network, Link Prediction, Data Mining, Privacy Protection Technique.

**Abstract:** Recently, social networking services have come to pervade every aspect of our lives, increasing the demand for the utilization of social network data. However, since the utilization of social network data has the risk of personal identification,  $k$ -anonymization methods have been proposed to protect privacy as a solution. In addition, the actual data contains missing values, which may reduce the utility of the data by applying conventional methods. In this paper, we propose a method for  $k$ -anonymization of social network data with link prediction, with the aim of improving the utility of anonymized data. Based on evaluation experiments on real data, we examine the utility of anonymized data.

## 1 INTRODUCTION

Recently, social networking services have come to pervade every aspect of our lives, increasing the demand for the utilization of social network data (Sivaganesan, 2021)(Rahmayani and Nofrialdi, 2022). However, since social network data contains personal information, the risk of personal identification arises (Liu and Terzi, 2008).

To take an example, data includes not only identifiers such as name, but also quasi-identifiers such as age and profession. Usually, the identifiers are removed when the data is provided to data providers. In this situation, however, there is a possibility that individuals can be identified from the personal information of quasi-identifiers provided, and sensitive personal information such as income and disease can be revealed. Therefore, it is necessary to consider the personal information when protecting privacy (Oishi, 2020).

In the case of social network data, structural information is also included such as friendships. For this reason, privacy-preserving methods focusing on structural information have been studied for anonymization of social network data.

In addition, the actual social network data may contain missing data. For example, Facebook, which


is one of the most influential social networking services, allows users to set friendships such to private. Therefore, when dealing with real data, it is effective to consider the missing friendships in the data and supplement links by link prediction (Zareie and Sakellariou, 2020). In this situation, it is important to suppress the information loss of anonymized data with link prediction.


In general, missing data is not considered during data anonymization. However, the presence of missing data increases the amount of information lost after anonymization, so the missing data should be considered when applying anonymization to real data.


This paper focuses on an anonymization method for social network data with link prediction, and aim to improve the utility of anonymized data.

In this study, we proposed a new indicator `k-maximum_degree` and an anonymization algorithm using the indicator for social network data with link prediction. In addition, we assumed that an attacker can estimate a node, and added an analysis algorithm for node estimation. Then, we conducted experiments adapting the proposed method to two actual social network data and evaluated the utility of the anonymized data. Experimental results showed that the anonymized data generated by the proposed method had the highest link match rate with the original data compared to those of other existing methods, indicating that it is superior in terms of data utility.

This paper is organized as follows. In Section 2, we describe the assumed environment in this study.

<sup>a</sup>  <https://orcid.org/0000-0002-2552-6717>

<sup>b</sup>  <https://orcid.org/0000-0002-1939-4455>

<sup>c</sup>  <https://orcid.org/0000-0001-6717-7028>

We explain some background knowledge related to this study in Section 3. Section 4 introduces the new indicator and algorithm for anonymization of social network data with link prediction. Then, we present the experiments and discuss the results in Section 5. Finally, we present the conclusions and the future work in Section 6.

## 2 ASSUMPTION

In this section, we show an assumed environment and an attack scenario in the situation. In this study, the notations used is shown in Table 1.

Table 1: Table of Notation.

$G(V, E)$	Graph $G$ with node set $V$ and edge set $E$
$G'(V', E')$	Graph $G'$ obtained by link prediction.
$ V  = n$	The number of nodes in $G$
$ E  = m$	The number of edges in $G$
$deg(v)$	The degree of node $v$ .
$k$	The anonymization requirement parameter.
$add$	The total number of links added by link prediction.
$add_i$	The number of links added by link prediction of $v_i$ .

### 2.1 Assumed Environment

In general, it is difficult for data analysts to voluntarily collect vast amounts of social network data, so it is very meaningful to analyze the data provided by data owners. In this paper, the case that data owners such as providers of social networking services provide the social network data to data analysts is assumed as shown in Figure 1.

Data owners know which links users have set as private, but no information related to these private links should be released to the outside. In case of such restrictions, it is necessary to anonymize social network data without using the private information. In this case, it is acceptable to infer private links only from public information (non-private links and nodes). Therefore, data owners may provide anonymized social network data after link prediction of public information.

In addition, the case where social network data contains missing links is assumed, and the following link prediction and anonymization is performed. The data flow is shown in Figure 2.

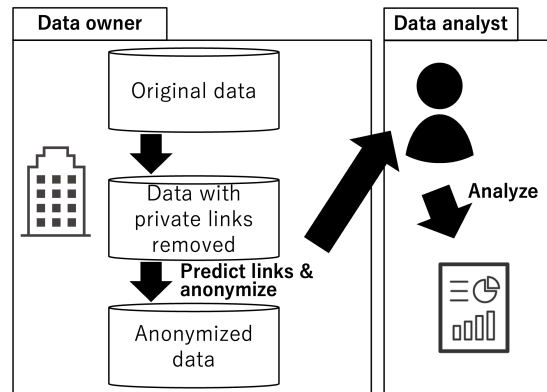


Figure 1: Assumed Environment.

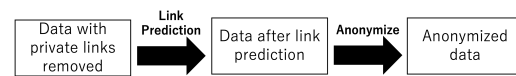


Figure 2: Data Flow.

### 2.2 Attacker's Background Knowledge

The following knowledge is set as the attacker's background knowledge.

- The degree of a node
- Total number of links added by link prediction

If the attacker knows the total number of links added by link prediction, the degree of node  $v$  will be at most the degree before link prediction plus the total number of links added.

$$deg(v) \leq deg'(v) \leq deg(v) + add \quad (1)$$

In this case, the attacker cannot determine which links have been added by link prediction.

## 3 BACKGROUND

### 3.1 Social Network Data

Social network data includes profiles such as name and profession, posts and friendships. Social network data consists of a set of nodes  $V$  and a set of edges  $E$ , which can be expressed as graph  $G = (V, E)$ . In this situation, the degree of a node indicates the number of friends.

### 3.2 Anonymization Techniques

Social network data contains personal privacy information. Therefore, data providers are required to pro-

protect the personal privacy information contained in the data before providing it.

k-anonymization (Sweeney, 2002) was proposed as one of the privacy protection methods. k-anonymization is a method to reduce the risk of re-identification to  $1/k$  by processing data so that there are  $k$  or more data with the same attributes. However, social network data includes structural information such as friendships as well as attribute information (Beigi and Liu, 2020). For this reason, it may not be safe to only anonymize attribute information.

Privacy preserving methods for network data can be divided into three main categories. The first method is graph modification. The approach is a method of anonymization by adding and deleting nodes and edges in the graph. The second method is generalization. The approach is a generalized method of grouping nodes and edges by clustering (Campan and Truta, 2008a)(Yu et al., 2017). The third method is uncertain graphs (Yan et al., 2018). The approach is based on uncertain graphs, which adds probabilities of existence to the edges of the graphs to add uncertainty to the graphs.

Among them, the k-anonymization method, which uses the degree of a vertex as a quasi-identifier, has been widely studied as graph modification approach (Casas-Roma et al., 2017a)(Casas-Roma et al., 2017b)(Casas-Roma et al., 2019).

A typical anonymity metric of social network data is k-degree (Liu and Terzi, 2008). k-degree guarantees the existence of at least  $k$  nodes with the same degree, and Liu et al. propose an anonymization method that uses k-degree to add or remove edges. k-degree is defined as follows.

**Definition 1.** An integer vector  $v$  satisfies  $k$ -anonymity if every value of  $v$  appears at least  $k$  times. Example:  $v = [5, 5, 3, 3, 2, 2, 2]$  satisfies 2-anonymity.

**Definition 2.** Let  $V$  be the degree sequence of the input graph and  $W$  be the degree sequence of the anonymized graph. In this case, Equation (2),  $v_i \in V$  and  $w_i \in W$  hold.

$$\Delta(V, W) = \sum_{i=1}^n |v_i - w_i| \quad (2)$$

For example, suppose that in the situation shown in Figure 3, an attacker has background knowledge of the degree of a particular user, and wants to identify a node of degree 2.

The degree of each node in Figure 3 is {'Ada':2, 'Bob':1, 'Cathy':1}. Since there is only one node with degree 2, Ada can be uniquely identified from the graph in Figure 3.

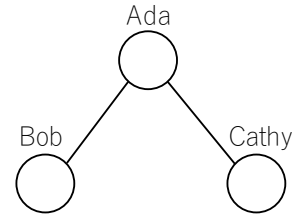


Figure 3: Social Network Graph.

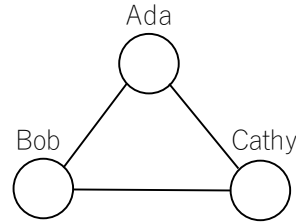


Figure 4: 3-degree Graph.

In Figure 4, the degree of each user becomes {'Ada':2, 'Bob':2, 'Cathy':2} by adding a new edge for Bob and Cathy. As a result, the risk of identification can reduce to  $1/3$ . At this point, we can say that the graph in Figure 4 satisfies 3-degree.

However, it is clear that there is a trade-off between data utility and anonymity in anonymization (Majeed and Lee, 2020). For the reason, it is necessary to apply anonymization methods with the needs of both data owners and data analysts depending on the characteristics of the data.

### 3.3 Problem in Link Prediction

In social network analysis, link prediction methods that use graph theory to predict the likelihood of the existence of links have been studied (Ahmad et al., 2020)(Campan and Truta, 2008b)(Lorrain and White, 1971)(Liben-Nowell and Kleinberg, 2003). Finding hidden relationships through link prediction is expected to increase the utility of social network data.

Common Neighbor and Centrality based Parameterized Algorithm (CCPA) was proposed in 2020. CCPA score is calculated based on common neighbors and centrality, and is defined as follows (Ahmad et al., 2020). In addition, the performance of the CCPA algorithm has been shown to be consistently better on some datasets than other algorithms (Sharma and Rai, 2022).

$$CCPA_{score} = \alpha \cdot (|\Gamma(u) \cap \Gamma(v)|) + (1 - \alpha) \cdot \frac{N}{d_{uv}} \quad (3)$$

In this case,  $\Gamma(u)$  and  $\Gamma(v)$  are the set of neighboring nodes of each node,  $N$  is the total number of nodes, and  $d_{uv}$  is the shortest distance between nodes.

The higher the score, the more likely it is that a link between nodes exists.

In this way, missing links in social network data can be predicted by link prediction. However, the question of an anonymization method for social network data with link prediction is still open in previous researches.

### 4 PROPOSED ALGORITHM

In this section, we propose k-maximum\_degree as an anonymity metric for social network data with link prediction, and risk of node re-identification as a possible attacker’s method of attack in the situation.

#### 4.1 Definition of k-maximum\_degree

In the situation of anonymized data after link prediction, we propose an anonymization indicator **k-maximum\_degree**.

**Definition 1.** A vector of integers  $V$  is  $k$ -anonymous, if every distinct value  $v_i \in V$  appears at least  $k$  times.

**Definition 2.** An undirected network  $G = (V, E)$  satisfies  $k$ -maximum\_degree anonymous, if the maximum degree in the degree sequence of  $G$  appears at least  $k$  times.

When there are at least  $k$  nodes with the maximum degree, the anonymized data after link prediction satisfies  $k$ -maximum\_degree. For example, if the degree sequence of  $G$  is  $[5, 5, 3, 2, 2, 2, 2, 1]$ , the graph  $G$  satisfies 2-maximum\_degree.

#### 4.2 Risk of Node Re-Identification

In this situation, the attack is assumed as shown in Figure 5.

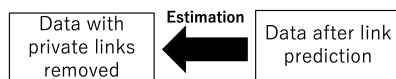


Figure 5: Data Attack Flow.

The degree of a node may be inferred from the total number of additions by link prediction. In this time, the risk of re-identification in a node is calculated by the Equation (4). The Equation (4) indicates the probability that a node is of a possible degree before link prediction.

$$deg\_probability = \frac{deg(v_i)C_{add_i} \times m - deg(v_i)C_{add-add_i}}{mC_{add}} \tag{4}$$

The total number of links added by link prediction is  $add$ , the total number of edges in graph is  $m$ , the degree of  $v_i$  is  $deg(v_i)$ , and the number of links added by link prediction of  $v_i$  is  $add_i$ .

For example, suppose we have a social network data with link prediction as shown in Figure 6.

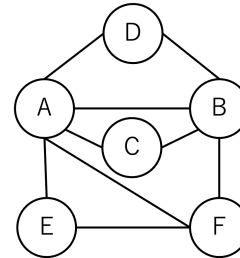


Figure 6: Social network data with link prediction.

This graph consists of 6 nodes and 9 edges, and the edges contain 2 additional links by link prediction. In this time, node  $A$  is the  $a$ -th node and the degree of  $A$  is 5 in the social network data with link prediction  $G'$ .

The probability that  $add_a = 0$  and  $deg(A) = 5$  is:

$$\frac{{}_5C_0 \times {}_{9-5}C_{2-0}}{{}_9C_2} = \frac{6}{36} = \frac{1}{6}$$

The probability that  $add_a = 1$  and  $deg(A) = 4$  is:

$$\frac{{}_5C_1 \times {}_{9-5}C_{2-1}}{{}_9C_2} = \frac{20}{36} = \frac{5}{9}$$

The probability that  $add_a = 2$  and  $deg(A) = 3$  is:

$$\frac{{}_5C_2 \times {}_{9-5}C_{2-2}}{{}_9C_2} = \frac{10}{36} = \frac{5}{18}$$

From the results, it is inferred that node  $A$  was likely of degree 4 before link prediction.

#### 4.3 Analysis Algorithm

Since it is assumed that attackers estimate the degree of users from the link prediction results and attempts to identify the nodes, an analysis algorithm estimates the pre-link prediction degree of each node from the total number of links added, which is one of the link prediction results.

An analysis algorithm that estimates the pre-link prediction degree of each node is shown in Algorithm 1.

Algorithm 1: Analysis Protocol.

---

```

1: Input:  $G'(V', E')$ ,  $add$ 
2: Output: Predicted degree sequence  $D_i(i = 0, \dots, n-1)$ 

3: function ESTIMATION_DEGREE( $G', add$ )
4:    $D \leftarrow [0] * n$ 
5:   for all  $i = 0, \dots, n-1$  do
6:      $max \leftarrow 0$ 
7:     for all  $s = 0, \dots, add-1$  do
8:       /* Calculate degree probability */
9:        $x_s \leftarrow calDegProbability(dev(v_i), s)$ 
10:      if  $x_s > max$  then
11:         $max \leftarrow x_s$ 
12:         $D_i \leftarrow dev(v_i) - s$ 
13:      end if
14:    end for
15:  end for
16:  return  $D_i(i = 0, \dots, n-1)$ 
17: end function

```

---

In  $calDegProbability(dev(v_i), s)$ , it receives the degree sequence  $deg(v_i)$  and the number of links added by link prediction  $s$  at  $v_i$ . Then, it estimates the probability that the degree before link prediction for  $v_i$  is  $dev(v_i) - s$ .

It is possible to find the most likely value as the degree before link prediction from the total number of links added by link prediction by using this analysis algorithm, and determine if a node with the degree satisfies k-anonymity.

#### 4.4 Anonymization Protocol

In the next, an anonymization algorithm that satisfies the k-maximum\_degree defined. In this algorithm, a method of adding edges for anonymization is adopted.

In the conventional method, the amount of modification of the graph is enormous, so the amount of modification is suppressed by adding and deleting graphs to maintain the quality of the anonymized graph. However, the proposed method is expected to suppress the amount of modification of graphs by anonymizing only the graph structure of some nodes, and to generate high-quality anonymized graphs even with only additions.

This anonymization algorithm is shown in Algorithm 2.

If any node has less than  $k$  nodes with the same degree in the predicted degree sequence  $D$ , then  $estimationDegreeAdd(D, list)$  is performed. In  $estimationDegreeAdd(D, list)$ , it receives a list  $list1$  of nodes with fewer than  $k$  nodes having the same degree in the predicted degree sequence  $D$ . Starting with

Algorithm 2: Anonymization Algorithm.

---

```

1: Input:  $G'(V', E')$ ,  $add, k$ 
2: Output: Anonymized graph  $G''$ 

3: function ANONYMIZATION( $G', add, k$ )
4:    $G'' \leftarrow$  empty graph
5:    $vd \leftarrow [0] * n$ 
6:    $D \leftarrow ESTIMATION\_DEGREE(G', add)$ 
7:    $list1 \leftarrow []$ 
8:   for all  $i = 0, \dots, n-1$  do
9:     if  $D.count(D[i]) < k$  then
10:       $list1.append(i)$ 
11:    end if
12:  end for
13:  if  $list1.size > 0$  then
14:     $vd \leftarrow estimationDegreeAdd(D, list1)$ 
15:  end if
16:   $vd \leftarrow degreeAnonymization(G', vd)$ 
17:   $G'' \leftarrow graphConstruction(G', vd)$ 
18:  /* Graph Construction */
19:  return  $G''$ 
20: end function

```

---

the smallest degree, add to each degree until there are at least  $k$  nodes with the same degree. When there are no more nodes with the same degree less than the  $k$  value, a list of changes for each node is returned.

In  $degreeAnonymization(G', vd)$ , it receives a graph  $G'$  obtained by link prediction and a list  $vd$  of changes for each node, and finds the minimum amount of additions required to have at least  $k$  nodes with k-maximum\_degree.

In  $graphConstruction(G', vd)$ , it receives a graph  $G'$  obtained by link prediction and a list  $vd$  of changes for each node. In addition, it adds edges so that the graph changes minimally and satisfies k-maximum\_degree.

## 5 EVALUATION EXPERIMENT

### 5.1 Experimental Setup

In this experiment, we evaluated the usefulness of anonymization methods for link prediction. In this experiment, we prepared missing social network data with 20% or 50% of edges removed, and anonymized the social network data using the following method.



- (a) Anonymization by k-degree
- (b) CCPA + k-degree
- (c) CCPA + k-maximum\_degree

In recent years, many anonymization methods for social network data have been studied, but k-degree algorithm, which uses only structural information and is widely used, was used for comparison.

In this study, k-degree and k-maximum\_degree were used for anonymization and CCPA was used for link prediction.

For link prediction, the CCPA parameter was set to the default value,  $\alpha = 0.8$ , and additional links were set as links with a normalized CCPA score of 0.9 or higher.

### 5.2 Evaluation Index

As an evaluation index, we used the link match rate. To evaluate the utility of degree-anonymous graph  $G_b(V, E_b)$ , the link match rate with the original data  $G(V, E)$  before edge removed was used. The link match rate is defined in the study as the edge intersection as  $EI(E, E_b) = |E_b \cap E| / |E|$  (Liu and Terzi, 2008). The higher the rate, the better the result, as it preserves the characteristics of the data.

### 5.3 Dataset

In the evaluation experiment in this study, we used two real social network datasets.

**Dolphins dataset** is used as the real data (Rossi and Ahmed, 2015). The dataset contains a social network of bottlenose dolphins, consisting of 62 nodes and 159 undirected edges, representing frequent associations among dolphins.

**Facebook dataset** is used as the real data (Leskovec and McAuley, 2012). The dataset includes profiles, circles, and ego networks. In the evaluation experiment, we used ego networks data composed of 4,039 nodes and 88,234 undirected edges as friendships.

### 5.4 Experimental Results

In the Dolphins dataset, the number of nodes decreased to **58** for the dataset with 20% edge removal, and to **56** for the dataset with 50% edge removal due to the effect of removing some of the links in the experiment.

As a result of link prediction, the dataset with 20% edge removal was increased to **132** edges with

**4** added, and the dataset with 50% edge removal was increased to **104** edges with **24** added.

The link match rates with the original data before edges removed are as shown in Table 2 and 3.

Table 2: The link match rate with the original data before edges removed (Dolphins: 20% removal).

	Recall↑	
	k=2	k=10
(a)k-degree	0.515	0.509
(b)CCPA+k-degree	0.507	0.512
(c)CCPA+k-maximum_degree	<b>0.977</b>	<b>0.872</b>

Table 3: The link match rate with the original data before edges removed (Dolphins: 50% removal).

	Recall↑	
	k=2	k=10
(a)k-degree	0.464	0.456
(b)CCPA+k-degree	0.364	0.347
(c)CCPA+k-maximum_degree	<b>0.840</b>	<b>0.720</b>

In the Facebook dataset, the number of nodes decreased to **4,018** for the dataset with 20% edge removal and to **3,944** for the dataset with 50% edge removal due to the effect of removing some of the links in the experiment. As a result of link prediction, the dataset with 20% edge removal was increased to **71,797** edges, with **1,209** added, and the dataset with 50% edge removal was increased to **63,956** edges, with **19,839** added.

The link match rates with the original data before edges removed were as shown in Table 4 and 5.

Table 4: The link match rate with the original data before edges removed (FB: 20% removal).

	Recall↑		
	k=2	k=5	k=10
(a)k-degree	NA	0.397	0.388
(b)CCPA+k-degree	NA	0.390	0.379
(c)CCPA+k-maximum_degree	<b>0.995</b>	<b>0.979</b>	<b>0.939</b>

Table 5: The link match rate with the original data before edges removed (FB: 50% removal).

	Recall↑		
	k=2	k=5	k=10
(a)k-degree	0.396	0.390	0.383
(b)CCPA+k-degree	0.275	0.271	0.267
(c)CCPA+k-maximum_degree	<b>0.899</b>	<b>0.888</b>	<b>0.863</b>

The conventional method sometimes failed in graph construction because it modifies the graph to minimize the amount of edge modification, and the anonymized data may not be obtained, as in the case of  $k = 2$  in (a) and (b) of Table 4.

The experimental results show that for (c) CCPA+k-maximum\_degree, the link match rate with

the original data before edges removal is larger than in (a) k-degree and (b) CCPA + k-degree.

In particular, in the case of Facebook dataset, the link match rate is larger for (c), (a), and (b), in that order, and the proposed method (c) outperforms the conventional method of k-degree anonymization in terms of the number of matched links.

In all results, the link match rate was worse for (b) CCPA+k-degree than for (a) k-degree. This may be due to an increase in the number of links in the dataset in the link prediction, and an increase in the amount of graph modification in anonymization, which added links that were not in the original data.

The anonymized data generated by the proposed anonymization method showed a higher link match rate than the other methods, suggesting that the characteristics of the data are better preserved.

There are two possible reasons for the high accuracy of the proposed method (c) CCPA+k-maximum\_degree.

First, the amount of graph modification can be reduced compared to anonymizing all nodes, because the nodes that need to be anonymized are selected based on the link prediction results.

Secondly, there is a difference in the graph modification method during anonymization. The proposed method (c) anonymizes the graph by adding links because the amount of link modification can be suppressed. On the other hand, the conventional methods (b) anonymizes by adding and deleting links in order to reduce the amount of link modification because the amount of link modification is enormous.

In Tables 4 and 5, the reason why NA was obtained for the results of conventional methods (a) and (b) is that the sum of the anonymized degree sequence is odd, and the anonymized graph cannot be constructed. In particular, conventional methods (a) and (b) require anonymization of degree sequences while minimizing the amount of modification so that all nodes satisfy k-degree, which may result in unfeasible degree sequences as graphs. In this respect, the proposed method (C) can suppress the number of nodes that are the main target of anonymization process and can easily generate a feasible degree sequence.

In addition, the results of the link match rate showed that as the value of the  $k$  value increases, the link match rates decreased. The number of links that match the original data before edge removal did not change much as the value of  $k$  increases, even though the number of edges increased. The reason why the number of links increases as the value of the  $k$  value increases is that the larger the  $k$  value, the higher the anonymity required, the more nodes need to be anonymized, and the amount of graph modification

increases.

In order to increase the utility of anonymized social network data with link prediction in the future, it is necessary to increase the number of links that match the original data before edge removal, such as by preferentially selecting links that are more likely to be present in the original data when adding links during anonymization.

## 6 CONCLUSIONS

In this paper, assuming that the real data contains missing data, we discussed an anonymity metric for anonymization of social network data with link prediction. We also applied the anonymization method for social network data with link prediction to real data and examined the utility.

The experimental results suggested that the proposed anonymization method preserves more data features than the application of conventional methods in anonymized data with link prediction.

In the future, we will conduct utility studies on other real data and safety evaluation, and investigate better ways to anonymize SNS data through link prediction.

## ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP21H03496, JP22K12157.

## REFERENCES

- Ahmad, I., Akhtar, M. U., Noor, S., and Shahnaz, A. (2020). Missing link prediction using common neighbor and centrality based parameterized algorithm. *Scientific reports*, 10(1):1–9.
- Beigi, G. and Liu, H. (2020). A survey on privacy in social media: Identification, mitigation, and applications. *ACM/IMS Trans. Data Sci.*, 1(1).
- Campan, A. and Truta, T. M. (2008a). A clustering approach for data and structural anonymity in social networks.
- Campan, A. and Truta, T. M. (2008b). Data and structural k-anonymity in social networks. In *International Workshop on Privacy, Security, and Trust in KDD*, pages 33–54. Springer.
- Casas-Roma, J., Herrera-Joancomartí, J., and Torra, V. (2017a). k-degree anonymity and edge selection: improving data utility in large networks. *Knowledge and Information Systems*, 50(2):447–474.

- Casas-Roma, J., Herrera-Joancomartí, J., and Torra, V. (2017b). A survey of graph-modification techniques for privacy-preserving on networks. *Artificial Intelligence Review*, 47(3):341–366.
- Casas-Roma, J., Salas, J., Malliaros, F. D., and Vazirgiannis, M. (2019). k-degree anonymity on directed networks. *Knowledge and Information Systems*, 61(3):1743–1768.
- Leskovec, J. and Mcauley, J. (2012). Learning to discover social circles in ego networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.
- Liben-Nowell, D. and Kleinberg, J. (2003). The link prediction problem for social networks. In *Proceedings of the twelfth international conference on Information and knowledge management*, pages 556–559.
- Liu, K. and Terzi, E. (2008). Towards identity anonymization on graphs. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 93–106.
- Lorrain, F. and White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, 1(1):49–80.
- Majeed, A. and Lee, S. (2020). Anonymization techniques for privacy preserving data publishing: A comprehensive survey. *IEEE access*, 9:8512–8545.
- Oishi, Keiichiro, e. a. (2020). Semantic diversity: Privacy considering distance between values of sensitive attribute. *Computers & Security*, 94:101823.
- Rahmayani, O. and Nofrialdi, R. (2022). The effect of utilization of social media instagram@ nanarfshop on buying interest of fisipol students university ekasakti padang. *Journal of Law, Politic and Humanities*, 2(2):85–94.
- Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *AAAI*.
- Sharma, A. and Rai, A. K. (2022). Analysis of link prediction algorithms based on similarity. In *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pages 1–6.
- Sivaganesan, D. (2021). Novel influence maximization algorithm for social network behavior management. *Journal of ISMAC*, 3(01):60–68.
- Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05):557–570.
- Yan, J., Zhang, L., Tian, Y., Wen, G., and Hu, J. (2018). An uncertain graph approach for preserving privacy in social networks based on important nodes. In *2018 International Conference on Networking and Network Applications (NaNA)*, pages 107–111.
- Yu, F., Chen, M., Yu, B., Li, W., Ma, L., and Gao, H. (2017). Privacy preservation based on clustering perturbation algorithm for social network. *Multimedia Tools and Applications*, 77:11241–11258.
- Zareie, A. and Sakellariou, R. (2020). Similarity-based link prediction in social networks using latent relationships between the users. *Scientific Reports*, 10(1):1–11.