International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhanced Privacy Preservation in Facebook Databases Using an Combined Approach with K-Member Fuzzy Clustering and Lyrebird Optimization Algorithm

Suresh R[1*], Dr. Rajavarman V N[2], Dr. Kevin Andrews S[3]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: Facebook's rapid growth has led to the collection of vast personal data, including age, location, occupation, and contact information, which poses significant privacy risks despite its utility in law enforcement and forensic investigations. The primary challenge is balancing the need for forensic access with protecting user privacy, ensuring that shared data does not allow for individual identification. Traditional anonymisation strategies like K-Anonymity (KA), l-diversity (LD), and t-closeness (TC) aim to safeguard personal data by removing or altering identifying information. However, these methods often prove inadequate, leaving data exposed to attribute and link disclosures, similarity attacks, and resulting in considerable information loss. This study introduces a more efficient anonymisation technique that combines K-member fuzzy clustering with the Modified Lyrebird Optimisation Algorithm (KFCMLOA). An Enhanced K-member version of the fuzzy c-means algorithm is first used to form balanced clusters, ensuring that each cluster has a minimum of K members. These clusters are then further refined and the data is anonymised using the Lyrebird Optimisation Algorithm (LOA). This discovery is important because it can protect anonymised Facebook databases from similarity attacks, identity, attribute, and link leaks, while reducing information loss.

Keyword: *Facebook, Lyrebird Optimization Algorithm, Anonymization Technique, Fuzzy Clustering Algorithm, Digital Forensics.*

1. Introduction

Facebook, one of the most widely used social networking sites globally [1], has grown rapidly in recent years and now contains a vast quantity of personal information, including names, email addresses, age, location, and occupation [2]. This wealth of information poses significant privacy risks. At the same time, law enforcement and forensic experts increasingly rely on social media data for investigations [3]. However, balancing forensic needs and user privacy is a critical challenge [4].

The main issue with sharing Facebook databases is the need to protect publicly released data from allowing individual identification. This includes safeguarding personal details such as age, location, job, names, email addresses, and other sensitive information [5]. Common approaches for safeguarding privacy involve anonymizing data by altering or eliminating specific details and utilizing methods such as KA [6], LD and TC [7]. However, these techniques have notable limitations, as they are susceptible to attribute and link disclosure [8], as well as similarity attacks. Furthermore, they frequently lead to significant information loss in the released datasets [9].

Therefore, there is a pressing need for more effective data anonymization methods [10]. This research's main goal is to develop a reliable method that significantly lowers the loss of Facebook data (such as age, location, occupation, interests, and email addresses) while protecting an anonymised database from similarity attacks and identity, attribute, and link disclosures [11].

In order to protect anonymised data [14], this study presents a hybrid anonymisation technique that combines the Modified Lyrebird Optimisation technique (KFCMLOA) [13] with K-member Fuzzy Clustering [12]. First, an Enhanced K-member version of fuzzy c-means clustering is used to produce balanced clusters, ensuring that each cluster has at least K members [15]. These clusters are then further refined using the LOA, which makes it possible to anonymise the data and network graph. Additionally, integration of opposition-based learning algorithms further strengthens the LOA performance [16], enhancing the overall robustness of the anonymization process.

This research is essential for safeguarding anonymized databases against various attacks [17], while effectively minimizing information loss in Facebook data.

The remainder of this research is structured as follows: Section II provides a summary of recent studies on online social networks in digital forensics. Explains the suggested approach, which combines the Modified Lyrebird

[1]*Research Scholar, Computer Science, Dr. M. G. R Educational and Research Institute, India. Email: sureshrmsscmphil@gmail.com*
[2]*Professor, Dr. M. G. R Educational and Research Institute, India. Email: nrajavarrman2003@gmail.com*
[3]*Professor, Department of Computer Applications, Dr. M. G. R Educational and Research Institute, India. Email: kevinnmca@gmail.com*

Optimisation process with an anonymisation process that makes use of K-member Fuzzy Clustering. In Section IV, we outline the experimental setup and the performance metrics used, followed by an in-depth analysis and discussion of the findings. Lastly, Section V concludes with a summary of the main points.

2. Literature Review

In 2020, Langari et al. [18] emphasized the difficulty of safeguarding social network data from individual identification during database sharing. Current anonymization methods, such as K-anonymity, frequently fall short in preventing different types of disclosures and lead to considerable information loss. This paper presents a novel anonymisation technique called the K-member Fuzzy Clustering and Firefly method (KFCFA) to overcome these problems. The outcomes on social network databases show how well KFCFA reduces data leaking.

In 2021, Zhang et al [19]. highlighted the growing popularity of social networks, where users share personal information like names, gender, and addresses. However, the large-scale collection and sharing of this data expose it to potential misuse by unauthorized parties. In order to address privacy concerns, a lot of research has been done on graph perturbation techniques, which change the local structure of social networks to maintain user anonymity. While effective, these techniques can reduce data usability by introducing random noise, necessitating a balance between privacy and usability. The study used five cutting-edge anonymisation algorithms on datasets from Facebook and Twitter and discovered that while most methods maintain some usefulness, no single technique works best across all datasets.

In 2023, Frimpong et al [20] addressed the privacy concerns of recommendation systems on online platforms, highlighting issues with user anonymity and data commercialization. They proposed RecGuard, a blockchain-based system to enhance user privacy and control. RecGuard features two smart contracts: RG-SH for handling user data and RG-ST for data storage, along with the incorporation of a graph convolutional network to identify malicious nodes. Their prototype demonstrated the effectiveness and privacy benefits of this approach.

In 2021, Jain et al [21] noted that online social networks have exploded in popularity due to fast-growing technology. This has led to numerous security and privacy issues from users sharing personal content. Attackers can maliciously use this information, especially targeting children. The study looks at a number of privacy and security risks, existing remedies, and defensive tactics to make online social networks more secure.

In 2023, Dehghani et al [22] introduced the LOA, a bio-inspired metaheuristic that imitates the protective reactions of lyrebirds to threats. The two phases of the LOA are exploitation, which serves as a concealment strategy, and exploration, which serves as an escape method. Its effectiveness was validated through twenty-two constrained optimization problems, where LOA consistently outperformed other algorithms.

3. Methodology

The study showed that obtaining KA minimises data leakage while identifying the best K-anonymous solution. To obtain approximate K-anonymous solutions, a variety of optimisation techniques have been used [23]. However, these methods have not adequately addressed the anonymity constraints, and there is a lack of strategies to minimize the distortion rate. To get over these limitations, we redefine the anonymity problem as a constrained multi-objective optimisation problem. This research introduces a solution that utilizes a hybrid approach merging a newly combined KFC method [24] and MLOA referred to as KFCMLOA. The overall workflow is illustrated in figure 1.

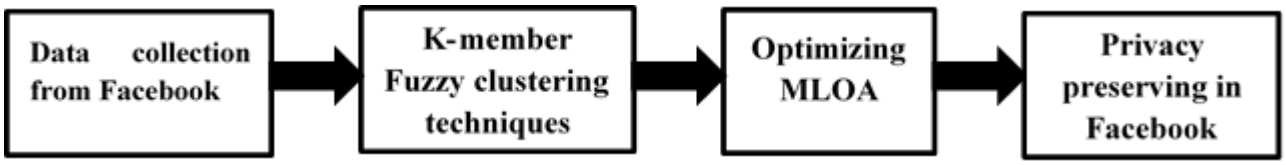

Fig 1: Proposed work flow

3.1. Databases

In this study, we utilized databases sourced from widely used social networks, particularly Facebook. Each database contains an attribute of personal attributes for individual users. Furthermore, an edge matrix is created based on the connections between users. The databases include information on 347 users, 224 data attributes, and 5,038 graph edges [18].

3.2. Enhanced K-member fuzzy clustering

Traditional clustering techniques encounter considerable challenges when applied to KA privacy-preserving contexts. Firstly, after the clustering process, some clusters may end up with no members. Furthermore, since some clusters may have a significantly larger number of members than K, these methods are inappropriate for clustering in KA-based applications. A novel approach called Enhanced KFC [25], which expands on Fuzzy C-

Means (FCM), has been put out to address these issues. With this approach, the clustering assignment vector and Cluster (C) parameters are set for the MLOA's initial population. The following describes KFC's clustering technique in depth.

3.2.1. Step by step process of KFC

Input: Initial Table T containing Dataset D & Graph E

Parameters: K & weights of attributes W

Output: Cluster & clustering assignment vector

Begin

Create the input matrix for clustering: CM = [E, W × D].

Determine an initial value for $Cluster = C_{initial}$ within the range of $[C_{min}, C_{max}]$.

Perform clustering CM into clusters using FCM.

Classify all clusters with at least K members as S_1 and the others as S_2.

For c = 1: Number of Clusters in S_2

For each cluster k = 1: C (k = c)

Calculate distance between the centroid of clusters c and k:Dist (c,k).

End For

Identify the closest cluster k to cluster c.

If cluster k belongs to S_2

Merge the two clusters k and c as a single cluster c.

Update the total number of clusters: C = C-1.

If the merged clusters contains at least K members

Eliminate the merged cluster from S_2 and add it into S_1.

End If

Else

Identify the K nearest members of cluster k and transfer them into cluster c.

Remove cluster c from S2 and add it into S_1.

End If

End For

Construct clustering assignment vector as the ratio of the number of clusters for each user to C.

End

3.3. Modified lyrebird optimization algorithm

MLOA is a metaheuristic algorithm that operates on a population basis. The starting solution in the proposed KFCMLOA technique is constructed using the structures for the number of clusters and the clustering assignment vector through KFC. Meanwhile the remaining structures namely, the feature selection vector, data modification matrix, and graph modification matrix are chosen at random. The algorithm proceeds through two important steps: updating the population and evaluating the objective function, until the maximum number of iterations (MaxIter) is reached. The best global solution found thus far is noted during each iteration. Upon completion of the MLOA, this recorded solution is considered the final output of the algorithm. Detailed explanations are provided here [18]. Figure 2 illustrate the Flowchart of Modified Lyrebird Optimization Algorithm.

3.3.1. Step by step process of MLOA

Start LOA

Input problem information: variables, objective function, and constraints.

Set LOA population size (P) and iterations (I).

Generate the initial population matrix at random, and Opposition based solution using Eq. (2): $y_{j,c} \leftarrow lb_c + r \cdot (ub_c - lb_c)$.

Evaluate the objective function.

Determine the best candidate solution.

For t = 1 to I

For i = 1 to P

Determine the type of lyrebird defense strategy against predator attack using Eq. (4). $Y_j \leftarrow \{based\ on\ Phase\ 1\ rn \leq 0.5\ based\ on\ Phase\ 2\ else$.

if $r_n \leq 0.5$ (choose Phase 1)

Determine candidate safe areas for the i-th lyrebird using Eq. (5): $SA_j \leftarrow \{Y_k, F_k < F_j\ and\ k \in \{1, 2, \ldots, P\}$.

Calculate the new position of the i-th LOA member using Eq. (6): $y_{j,i}^{N1} \leftarrow y_{j,i} + r_{j,i} \cdot (SSA_{j,i} - D_{j,i} \cdot y_{j,i})$.

Update the i-th LOA member using Eq. (7): $Y_i \leftarrow \{Y_j^{N1}, F_j^{N1} < F_i\ Y_j, else$

else (choose Phase 2)

Calculate the new position of the i-th LOA member using Eq. (8): $y_{j,i}^{N2} \leftarrow y_{j,i} + \left(1 - 2r_{j,i}\right) \cdot \frac{ub_i - lb_i}{t}$.

Update the i-th LOA member using Eq. (9): $X_i \leftarrow \{Y_j^{N2}, F_j^{N2} < F_j\ Y_j, else$

end (if)

end (For i = 1 to P

Save the best candidate solution so far.

end (For t = 1 to I)

Output the best quasi-optimal solution obtained with the LOA.

End LOA

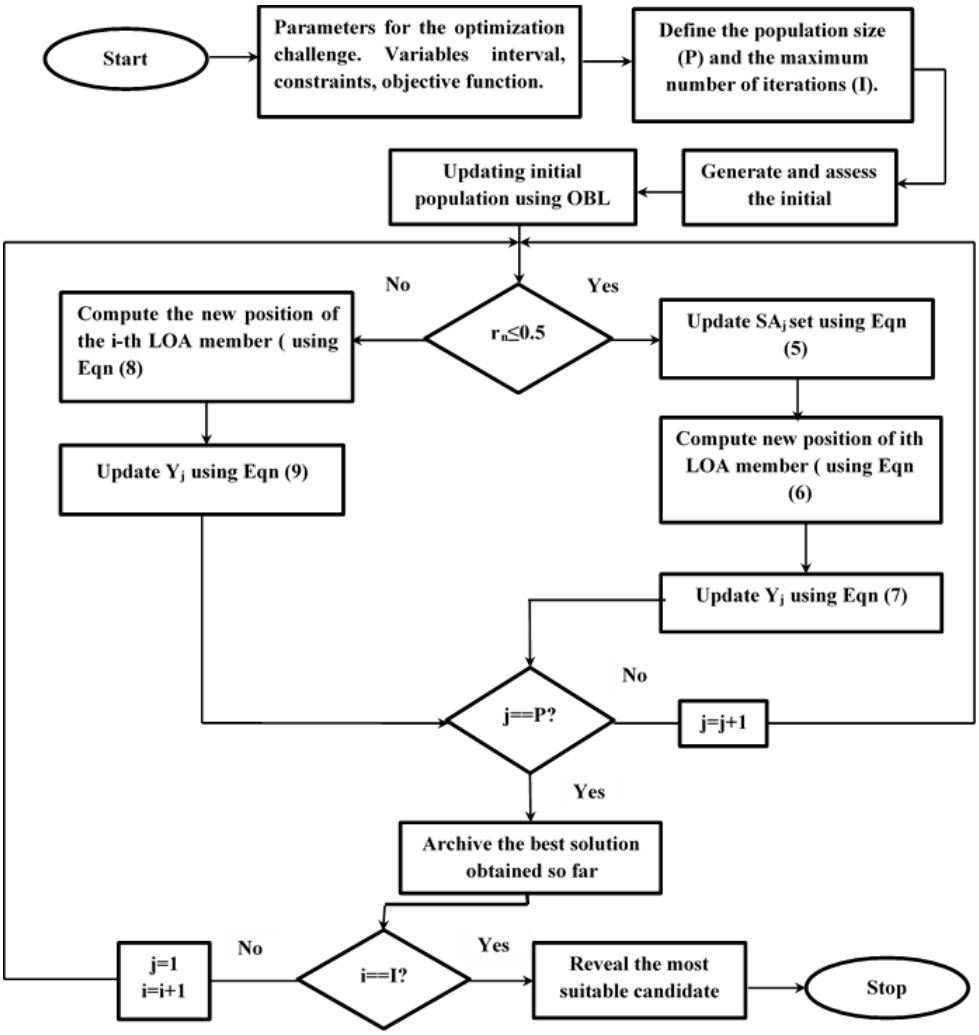

Fig 2: Flowchart

3.4. Evaluation of the objective function

In order to maximise anonymity performance during clustering and satisfy the constraints of KA, LD, and TC, the anonymity problem is presented as a constrained optimisation challenge. For this, a multi-objective function is constructed with three primary goals and three constraints. In order to optimise cluster balance, the goals are to decrease clustering error, minimise the cernability avg index, and minimise the average distortion ratio of the modified table caused by changes in features, data, and graphs. Additionally, to ensure that the three constraints—KA, LD, and TC—are met for each cluster, a penalty function that quantifies the number of unmet requirements is supplied. The restricted multi-objective function, which consists of a cost function and a penalty function, is represented by Eq. (1).

$$Objective\ function = Cost\ \ function \times (1 + Penalty\ function) \quad (1)$$

The cost function includes three objectives that need to be reduced. Eq. (2) explains how these goals are merged into a single-objective function using a weighted average.

$$Cost = w_{\text{Clustering Error}} \times \text{Clustering Error} + w_{\text{Cernability AVG index}} \times \text{Cernability AVG index} + w_{\text{AverageDistortion Rate}} \times \text{Average Distortion Rate} \quad (2)$$

The weight s$w_{CClustering\ ErrorE}$, $w_{Cernability\ AVG\ index}$ and $w_{Average\ Distortion\ Rate}$ $(w_{Clustering\ Error} + w_{Cernability\ AVG\ index} + w_{AAverage\ Distortion\ RateDR} = 1)$ are fixed parameters that determine the relative importance of the clustering error, cernability average index, and average distortion rate objectives. A larger weight indicates a greater influence of that objective on the overall Cost. The three objectives can be defined in the following equations:

$$\text{Clustering Error} = \frac{IntraDist}{IntraDist} = \frac{\frac{1}{N}\sum_{i=1}^{N} d\left(note_i, center_{CLA(i)}\right)}{\frac{1}{C(C-1)}\sum_{k=1}^{C}\sum_{k=1}^{C} d\left(center_j, center_k\right)} \quad (3)$$

$$\text{Cernability AVG index} = \frac{\left(\frac{N}{C}\right)}{K} \quad (4)$$

$$\text{Average Distortion Rate} = \frac{DR_F + DR_D + DR_E}{3} \quad (5)$$

With d $d(node_I, center_{CLA(i)})$ representing the sample i's Euclidean distance to its cluster centroid and d $d(center_J, center_K)$ representing the distance between

the centre of cluster j and the centre of cluster k, Eq. (3) provides the clustering error (where 0 < clustering error < 1) as the ratio of the average intra-cluster distances to the average inter-cluster distances. In Eq. (4), the Cernability avg index (cernability avg index ≥ 1) indicates the degree of cluster balancing; the lower the Cernability avg index, the more cluster balancing there is.

The proposed KFCMLOA satisfies the KA requirement by grouping users into K-member clusters, each of which comprises at least K users. As a result, each user can't be distinguished from at least K-1 other users, offering 1/K defence against identity theft. Furthermore, until all clusters satisfy the LD and TC requirements, the values of the sensitive characteristics are randomly modified. The LD requirement is met if the sensitive attributes of the users in a cluster have at least L distinct values. The TC requirement is also met if the discrepancy between the global data and the distribution of sensitive attributes within each cluster is less than I.

The KFCMLOA framework incorporates a penalty function into the issue formulation to address the KA, LD, and TC restrictions. According to Eq. (6), the penalty function is determined by the number of unfulfilled requirements, namely the number of breaches of KA, LD, and TC constraints over all clusters. If the KA holds for cluster i, $K-not(i) = 0$, otherwise, $K\_not(i) = 1$. similarly, if LD and/or TC has not been satisfied for cluster i, $L\_not(i) = 1\ and/or\ T\_not(i) = 1$, respectively.

$$Penalty = \text{Penalty of KA} + \text{Penalty of LD} + \text{Penalty of TC} \quad (6)$$

$$\text{Penalty of KA} = \sum_{i=1}^{C} K\_not(i) \quad (7)$$

$$\text{Penalty of LD} = \sum_{i=1}^{C} L\_not(i) \quad (8)$$

$$\text{Penalty of TC} = \sum_{i=1}^{C} T\_not(i) \quad (9)$$

3.5. Performance analysis

3.5.1. Simulation setup

The KFCMLOA method was successfully implemented on a personal computer with a 2.6 GHz Core i7 processor and 16 GB of RAM using MATLAB R2018b. The effectiveness of KFCMLOA was assessed by contrasting its outcomes with those of four clustering-based anonymity methods: T-closeness L-diversity KA, P-sensitive KA (PSKA), K-member fuzzy KA (KMFKA), and K-anonymity 3 Layers (TCLK3L).

3.5.2. Simulation results

The proposed KFCMLOA is compared with MLOA, FCMLOA, KA, PSKA, KMFKA, and TCLK3L to assess the impact of the various concepts employed. The following part displays the simulation results for a number of scenarios while accounting for different values for the system model parameters (K, L, and T). As previously stated, increasing K and L while decreasing T enhances defence against various attacks, but it also makes the problem more complex. The suggested algorithms use a penalty function to ensure adherence to $KA, LD, and\ TC$ restrictions for every combination of the system parameters $K, L, and\ T$. To assess the effectiveness of various anonymity techniques at various privacy levels, six scenarios were created:

- *Scenario 1* (simple): K = 4, L = 3, T = 0.7.
- *Scenario 2* (medium): K = 6, L = 4, T = 0.5.
- *Scenario 3* (hard): K = 8, L = 6, T = 0.3.

3.5.3. Simulation findings for scenario 1

In the first instance, K = 4, L = 3, and T = 0.7 were the three anonymity parameters. Each strategy was assessed across ten consecutive runs due to the stochastic nature of the employed approaches. Table 1 displays the findings for the Facebook database, including the average and standard deviation for the ten runs. While PSKA satisfies both the KA and LD criteria, both the KA and KMFKA procedures meet the KA condition, as seen in table 1.

However, when combined with the three recommended methods, TCLK3L ensures that the KA, LD, and TC requirements are met simultaneously. KMFKA yielded the lowest error rates in clustering error, cernability average index, and average distortion ratio as expected, but it is not the optimal method because it does not meet the LD and TC criteria, which leads to a high penalty function. Only TCLK3L, which takes into account all three limitations, is comparable to the suggested approaches out of the four methods that are currently in use in the literature. Reducing the clusters to 14 resulted in a notable rise in the cernability average index to 6.19, which are around three to four times higher than our techniques. The TCLK3L algorithm used a big number of clusters to obtain zero penalties. Furthermore, the TCLK3L algorithm's average distortion ratio significantly outperforms the suggested techniques, and its clustering error is around double that of our clustering mistakes. To sum up, the TCLK3L algorithm may simultaneously satisfy the KA, LD, and TC requirements, but at the cost of notable error rates in the clustering error, cernability average index, and average distortion ratio.

Table 1: Evaluation of different methods in scenario 1 for the facebook dataset.

No of Parameters	KA	PSKA	KMFKA	TCLK3L	MLOA	FCMLOA	KFCMLOA
Cluster	21.2 ± 1.8	17.9 ± 2.2	86 ± 0	14 ± 1.5	43.8 ± 3.7	47.1 ± 2.4	48.4 ± 2.3
Clustering error	0.081 ± 0.007	0.118 ± 0.003	0.017 ± 0.001	0.19 ± 0.008	0.088 ± 0.0028	0.071 ± 0.002	0.068 ± 0.0015
Cernability avg index	4.08 ± 0.38	4.83 ± 0.52	1.001 ± 0	6.19 ± 0.64	1.98 ± 0.123	1.84 ± 0.085	1.79 ± 0.082
Average distortion ratio	0 ± 0	0.001 ± 0.0001	0 ± 0	0.07 ± 0.0031	0.004 ± 0.0002	0.024 ± 0.0007	0.005 ± 0.0001
Penalty of KA	0	0	0	0	0	0	0
Penalty of LD	2.3 ± 0.41	0	23.6 ± 4.5	0	0	0	0
Penalty of TC	22.7 ± 3	17.4 ± 2.3	71.5 ± 5.1	0	0	0	0
Cost function	0.457 ± 0.019	0.554 ± 0.025	0.111 ± 0.005	0.758 ± 0.031	0.257 ± 0.011	0.225 ± 0.009	0.214 ± 0.008
Penalty function	26 ± 3.4	17.4 ± 2.3	95.1 ± 9.6	0	0	0	0
Objective function	12.36 ± 0.66	10.17 ± 0.61	10.7 ± 0.45	0.758 ± 0.031	0.257 ± 0.011	0.225 ± 0.009	0.214 ± 0.008
Time (s)	0.7	3.1	3.4	2.3	43	45	46

In contrast to the TCLK3L approach, the suggested KFCMLOA algorithm significantly lowers clustering error, cernability average index, and average distortion ratio while producing balanced clusters, according to the data in Table 1. This increase is mostly due to the algorithm's ability to structure the problem as a restricted multi-objective optimisation problem and solve it using a fuzzy-metaheuristic technique. However, it is important to acknowledge that the implementation of a population-based metaheuristic algorithm lengthens the execution timeframes of the proposed methodologies. Although the KFCMLOA technique has a slightly higher average distortion ratio than the FCFMLOA and MLOA methods, it performs better for the clustering error and cernability avg index criterion. Overall, the objective function values for the TCLK3L, MLOA, FCMLOA, and KFCMLOA algorithms, each attaining a 0 penalty, are 0.758, 0.257, 0.225, and 0.214, respectively. As a result, the KFCMLOA algorithm emerges as the most effective method in the first experiment.

Ultimately, a comparison of various error functions Facebook databases using TCLK3L, FMLO, FCMLOA, and KFCMLOA, specifically for clustering error, cernability avg index, average distortion ratio, and objective function. These algorithms are chosen because they consistently ensure that all constraints are satisfied, indicated by a penalty of zero. In conclusion, as compared to TCLK3L, MLOA, and FCMLOA, the KFCMLOA algorithm reduces the average total objective function Facebook databases by 51%, 13%, and 5%, respectively.

3.5.4. Simulation findings for scenario 2

To overcome the limitations of the anonymity procedure, the experiment's K, L, and T parameters were changed. Specifically, raising K and L while decreasing T produced K = 6, L = 4, and T = 0.5. Table 2 summarises the outcomes of the different approaches, averaged across ten iterations. The table 2 shows that in this experiment, the KA, PSKA, and KMFKA approaches exhibit a non-zero penalty function. These techniques produce a notably high Objective Function, the main evaluation criterion, even if they come at a fair price. Similar to the first experiment, all KA, LD, and TC criteria are simultaneously satisfied by the three proposed approaches and the TCLK3L strategy, resulting in a penalty function of zero. Thus, the cernability average index criterion of the TCLK3L method is 5.21, which is approximately four times higher than that of the KFCMLOA algorithm. Additionally, TCLK3L's average distortion ratio and clustering error are roughly

two and three times higher than KFCMLOA's, respectively. The suggested KFCMLOA algorithm performed the best, producing objective function values of 0.69, 0.196, 0.174, and 0.168 for TCLK3L, MLOA, FCMLOA, and KFCMLOA, respectively, in line with Scenario 1. On average, KFCMLOA reduced the objective function by 55%, 12%, and 1.5% compared to TCLK3L, MLOA, and FCMLOA, respectively, across these databases.

Table 2: Evaluation of different methods in scenario 2 for the facebook dataset

No of Parameters	KA	PSKA	KMFKA	TCLK3L	MLOA	FCMLOA	KFCMLOA
Cluster	18.8	16.2	57	11.2	42.6	47.5	48.7
Clustering error	0.098	0.135	0.03	0.265	0.085	0.072	0.07
Cernability avg index	3.07	3.58	1.014	5.21	1.37	1.24	1.2
Average distortion ratio	0	0.001	0	0.046	0.031	0.024	0.021
Penalty of KA	0	0	0	0	0	0	0
Penalty of LD	2.5	0	18.3	0	0	0	0
Penalty of TC	19.8	17	36.2	0	0	0	0
Cost function	0.363	0.442	0.12	0.69	0.196	0.174	0.168
Penalty function	22.3	17	54.5	0	0	0	0
Objective function	8.36	7.97	6.6	0.69	0.196	0.174	0.168
Time (s)	0.4	2.9	1.6	1.8	42	43	45

3.5.5. Simulation findings for scenario 3

The limits imposed by the parameters K, L, and T have become more apparent in this scenario with values set at K = 8, L = 6, and T = 0.3. Each method underwent ten consecutive tests, and the results from these methods are compiled in Table 3. Remarkably, the TCLK3L algorithm yielded only 4.1 clusters, resulting in a cernability average index of 10.8, which is quite high. Additionally, the computational requirements of TCLK3L are roughly 6 to 7 times higher than those of the proposed KFCMLOA. The objective function values for TCLK3L, MLOA, FCMLOA, and KFCMLOA are 1.79, 0.332, 0.258, and 0.239, respectively. On average, the KFCMLOA algorithm demonstrated a reduction in the objective function of 63%, 10%, and 5% compared to TCLK3L, MLOA, and FCMLOA, respectively. Thus, in line with previous scenarios, KFCMLOA emerges as the most effective algorithm for Scenario 3.

Table 3: Evaluation of different methods in scenario 3 for the facebook dataset

No of Parameters	KA	PSKA	KMFKA	TCLK3L	MLOA	FCMLOA	KFCMLOA
Cluster	13.4	10.3	43	4.1	23.2	28.7	30.4
Clustering error	0.161	0.289	0.045	1.138	0.233	0.17	0.15
Cernability avg index	3.33	4.33	1.008	10.8	1.87	1.5	1.41
Average distortion ratio	0	0.002	0	0.091	0.139	0.022	0.027
Penalty of KA	0	0	0	0	0	0	0
Penalty of LD	2	0	21	0	0	0	0
Penalty of TC	21	12	73	0	0	0	0
Cost function	0.43	0.6	0.12	1.79	0.332	0.258	0.239
Penalty function	23	12	94	0	0	0	0
Objective function	10.32	7.9	12.1	1.79	0.332	0.258	0.239
Time (s)	0.3	9.5	0.9	2.2	33	37	37

4. Discussion

In three different cases, the simulation results for the suggested methodology are shown in this part along with a comparison with current approaches. The results show that the KFCMLOA and TCLK3L algorithms successfully satisfy the requirements for KA, LD, and TC all at once. Consequently, these algorithms are adept at protecting the published database from similarity, link, attribute, and

identity disclosure assaults. However, because the TCLK3L algorithm ignores faults related to the anonymity process, the error metrics—clustering error, discernibility average index, average distortion ratio, and objective function—are around three times higher for the TCLK3L algorithm compared to the other approaches.

In the new dataset, the KFCMLOA algorithm shows a better capacity to reduce the discernibility average index, the average distortion ratio, and clustering errors while producing more balanced clusters. Three main factors are responsible for this improvement: (1) the problem was successfully formulated as a constrained multi-objective optimisation challenge that involves both anonymisation and clustering; (2) a K-member balanced fuzzy clustering technique was used; and (3) an efficient solution representation was provided within the MLOA framework. Simulation results demonstrate that the KFCMLOA algorithm outperforms the MLOA technique in terms of both performance and convergence speed since K-member fuzzy clustering is integrated into the MLOA methodology.

5. Conclusion

This research presents the KFCMLOA methodology designed to anonymize Facebook data and enhance user privacy. By employing the K-member fuzzy clustering approach, it generates balanced clusters that are further refined using the MLOA. This optimization aims to minimize errors in clustering and reduce information loss while ensuring privacy is maintained. Simulation results on Facebook data indicate that KFCMLOA surpasses existing methods in decreasing information loss and clustering errors, though it requires longer execution times due to its iterative nature. In summary, the KFCMLOA algorithm successfully safeguards against identity, attribute, and link disclosures, along with similarity attacks, establishing itself as a strong solution for anonymizing Facebook data while ensuring user privacy is maintained.

Declaration of Conflicting Interests

The authors have affirmed the absence of any potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

The authors did not receive any financial support for the research, authorship, and/or publication of this article.

Reference

[1] Tønnesson, S., Zaw Oo, M. and Aung, N. L., "Pretending to be States: The use of Facebook by Armed Groups in Myanmar". Journal of Contemporary Asia, 52(2), pp. 200-225, (2022). https://doi.org/10.1016/j.jretconser.2021.102501

[2] Zhu, T., Li, J., Hu, X., Xiong, P. and Zhou, W., "The Dynamic Privacy-Preserving Mechanisms for Online Dynamic Social Networks". IEEE Transactions on Knowledge Data Engineering, 34(6), pp. 2962-2974, (2020). DOI: 10.1109/TKDE.2020.3015835

[3] Powell, T. and Haynes, C., "Social Media Data in Digital Forensics Investigations". Digital Forensic Education: An Experiential Learning Approach, pp. 281-303, (2020). https://doi.org/10.1007/978-3-030-23547-5_14

[4] Stoyanova, M., Nikoloudakis, Y., Panagiotakis, S., Pallis, E. and Markakis, E. K., "A Survey on the Internet of Things (Iot) Forensics: Challenges, Approaches, and Open Issues". IEEE Communication Survey & Tutorials, 22(2), pp. 1191-1221, (2020). DOI: 10.1109/COMST.2019.2962586

[5] Quan-Haase, A. and Ho, D., "Online Privacy Concerns and Privacy Protection Strategies Among Older Adults in East York, Canada". Journal of the Association for Information Science and Technology, 71(9), pp. 1089-1102, (2020). https://doi.org/10.1002/asi.24364

[6] Gangarde, R., Sharma, A., Pawar, A., Joshi, R. and Gonge, S., "Privacy Preservation in Online Social Networks Using Multiple-Graph-Properties-Based Clustering to Ensure K-Anonymity, L-Diversity, and T-Closeness". Electronics, 10(22), pp. 2877, (2021). https://doi.org/10.3390/electronics10222877

[7] Mehta, B. B. and Rao, U. P., "Improved L-Diversity: Scalable Anonymization Approach for Privacy Preserving Big Data Publishing". Journal of King Saud University- Computer and Information Science, 34(4), pp. 1423-1430, (2022). https://doi.org/10.1016/j.jksuci.2019.08.006

[8] Chu, Z., He, J., Li, J., Wang, Q., Zhang, X. and Zhu, N., "SSKM_DP: Differential Privacy Data Publishing Method Via SFLA-Kohonen Network". Applied Science, 13(6), pp. 3823, (2023). https://doi.org/10.3390/app13063823

[9] Zigomitros, A., Casino, F., Solanas, A. and Patsakis, C., "A Survey on Privacy Properties for Data Publishing of Relational Data". IEEE Access, 8, pp. 51071-51099, (2020). DOI: 10.1109/ACCESS.2020.2980235

[10] Majeed, A. and Lee, S., "Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey". IEEE Access, 9, pp. 8512-8545, (2020). DOI: 10.1109/ACCESS.2020.3045700

[11] Torra, V. and Navarro-Arribas, G., "Attribute Disclosure Risk for K-Anonymity: The Case of Numerical Data". International Journal of Information Security, 22(6), pp. 2015-2024, (2023). https://doi.org/10.1007/s10207-023-00730-x

[12] Lawrance, J. U., Jesudhasan, J. V. N. and Thampiraj Rittammal, J. B., "Parallel Fuzzy C-Means Clustering

Based Big Data Anonymization Using Hadoop Mapreduce". Wireless Personal Communications, pp. 1-28, (2024). https://doi.org/10.1007/s11277-024-11101-7

[13] Narwal, A., "Resource Utilization Based on Hybrid WOA-LOA Optimization With Credit Based Resource Aware Load Balancing and Scheduling Algorithm for Cloud Computing". Journal of Grid Computer, 22(3), pp. 1-24, (2024). https://doi.org/10.1007/s10723-024-09776-0

[14] Majeed, A. and Lee, S., "Attribute Susceptibility and Entropy Based Data Anonymization To Improve Users Community Privacy and Utility in Publishing Data". Applied Intelligent, 50(8), pp. 2555-2574, (2020). https://doi.org/10.1007/s10489-020-01656-w

[15] Ikotun, A. M. and Ezugwu, A. E., "Enhanced Firefly-K-Means Clustering with Adaptive Mutation and Central Limit Theorem for Automatic Clustering of High-Dimensional Datasets". Applied Science, 12(23), pp. 12275, (2022). https://doi.org/10.3390/app122312275

[16] Ewees, A. A., Elaziz, M. A. and Oliva, D., "A New Multi-Objective Optimization Algorithm Combined With Opposition-Based Learning". Expert Systems with Applications, 165, pp. 113844, (2021). https://doi.org/10.1016/j.eswa.2020.113844

[17] Khan, R., Tao, X., Anjum, A., Sajjad, H., Malik, S. U. R., Khan, A. and Amiri, F., "Privacy Preserving for Multiple Sensitive Attributes Against Fingerprint Correlation Attack Satisfying C-Diversity". Wireless Communications and Mobile Computing 2020, 1, pp. 8416823, (2020). https://doi.org/10.1155/2020/8416823

[18] Langari, R. K., Sardar, S., Abdollah, S., Mousavi, A. and Radfar, R., "Combined Fuzzy Clustering and Firefly Algorithm for Privacy Preserving in Social Networksm". Expert Systems Applications, 141, pp. 112968, (2020). https://doi.org/10.1016/j.eswa.2019.112968

[19] Zhang, C., Jiang, H., Cheng, X., Zhao, F., Cai, Z. and Tian, Z., "Utility Analysis on Privacy-Preservation Algorithms for Online Social Networks: An Empirical Study". Personal and Ubiquitous Computing, 25, pp. 1063-1079, (2021). https://doi.org/10.1007/s00779-019-01287-0

[20] Frimpong, S. A., Han, M., Boahen, E. K., Sosu, R. N. A., Hanson, I., Larbi-Siaw, O. and Senkyire, I. B., "RecGuard: An Efficient Privacy Preservation Blockchain-Based System for Online Social Network Users". Blockchain: Research Applications, 4(1), pp. 100111, (2023). https://doi.org/10.1016/j.bcra.2022.100111

[21] Jain, A. K., Sahoo, S. R. and Kaubiyal, J., "Online Social Networks Security and Privacy: Comprehensive Review and Analysis". Complex Intelligent Systems, 7(5), pp. 2157-2177, (2021). https://doi.org/10.1007/s40747-021-00409-7

[22] Dehghani, M., Bektemyssova, G., Montazeri, Z., Shaikemelev, G., Malik, O. P. and Dhiman, G., "Lyrebird Optimization Algorithm: A New Bio-Inspired Metaheuristic Algorithm for Solving Optimization Problems," Biomimetics, 8(6), pp. 507, 2023. https://doi.org/10.3390/biomimetics8060507

[23] Canbay, Y., "On the Complexity of Optimal K-Anonymity: A New Proof Based on Graph Coloring". IEEE Access, (2024). DOI: 10.1109/ACCESS.2024.3424399

[24] Yang, Y., Li, M. and Ma, X., "A Point Cloud Simplification Method Based On Modified Fuzzy C-Means Clustering Algorithm with Feature Information Reserved". Mathematical Problems in Engineering 2020, 1, pp. 5713137, (2020). https://doi.org/10.1155/2020/5713137

[25] Subramaniam, M., Kathirvel, A., Sabitha, E. and Basha, H. A., "Modified Firefly Algorithm and Fuzzy C-Mean Clustering Based Semantic Information Retrieval". Journal of Web Engineering, 20(1), pp. 33-52, (2021). DOI: 10.13052/jwe1540-9589.2012