International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

K-Anonymity Enhancement for Privacy Preservation with Hybridization of Cuckoo Search and Neural Network using Clustering

Inderjit Kaur, Vijay Bhardwaj

Abstract*: *Expansion of social network and the publication of its data have directed the risk of disclosure of individuals' confidential information. Privacy preservation is a must thing before service providers publish the network data. In recent years, privacy in social network data has become the most concerned issue as it has gripped our lives in a dramatic manner. Numerous anonymization methods are there that assists in privacy preservation of social networking and among all, k-anonymity is the utmost one that helps in providing the security by developing graph and nodes degree. In this manuscript, the enhancement of K-anonymity has been addressed with major changes in node editing methodology. The clusters are developed with the integration of the same degree in one group and the procedure is iterated till the identification of noisy data. An advanced Cuckoo Search is commenced for minimizing the node miss placement in groups. The results of the Cuckoo Search are integrated with Feed Forward Back Propagation Neural Networks to cross-check the structure and to reduce the node miss placement in groups. Average Path Length (APL) and Information parameters are measured for the evaluation and comparative analysis and the effectiveness of the research has been checked by comparing the results of Aanchal Sharma and P. R. Bhaladhare. There is a diminution of 14.6% while comparing APL with Aanchal Sharma and 8.61% and 10.38% of reduction is shown with Aanchal Sharma and P. R. Bhaladhare for Information loss.

Keywords*: *Preservation, K-anonymity, Clustering, Cuckoo search, Neural network, APL, Information loss

I. INTRODUCTION

Considering the numerous big data resources, social networking has put in substantial data amount with each characteristic of front end and backend [1]. 1.65 billion users are covered by Facebook having 1 billion as active user of month, 600 million users are covered by twitter having 0.5 billion tweets every day, 304 million users are covered by Amazon having 9.65 billion items operated every year, 829 million active and 210 million concurrent users are covered by Tencent QQ and so on. So, with such a huge variety and scale of data, the analysis of social networking is now a significant task for the classification of end users, prediction of buying interest with the divination of event incidents and so on[2].

Manuscript published on 30 August 2019.
*****Correspondence Author(s)
Inderjit Kaur, Computer Applications, Guru Kashi University Talwandi Sabo, Punjab, India.
Dr. Vijay Bhardwaj, Computer Applications, Guru Kashi University Talwandi Sabo, Punjab, India.

Even though, the profuse social data has varied advantages, that it can enhance the rigid privacy apprehensions as well. Every user concerned with social networking is usually linked with an attribute set with sensitive attributes, such as sexual orientation, gender and location. There is a possibility that the personal information can be suppressed via third parties, such as social media, marketer and analysts. Some third party with the malevolent object is considered as Adversaries and they may violate the privacy of user with the integration of sensitive data initially [3]. The user these days are concerned more about the privacy and has now has become conventional for publishing the sensitive and the personal data that leads to degradation of data publishing scale and can force the users for publishing the anonymised data. Consequently, the variance amongst the data utility and the privacy concern has promoted the adversaries for exploiting the sensitive information inside the published data [4].

A. Motivation illustration

Having a little knowledge for an individual vertex in social networking, some victims could be attacked by the adversaries. For instance, a synthesized social network has been considered as friends as depicted in Fig.1. Every vertex in a network illustrates a user [5].

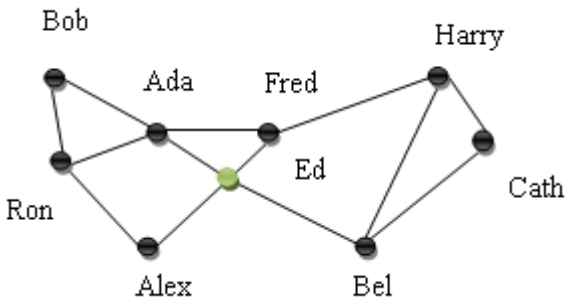

Fig.1Social Network

To publish the data to the network, privacy becomes the hot topic. So, for privacy preservation, it is vital to remove each identity as depicted in Fig.2.

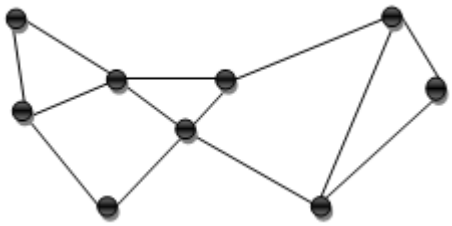

Fig.2Anonymous node network

There is a possibility that an adversary with the information of an individual's neighbour can leak the privacy. What if an adversary recognizes that Ed has two friends with each other recognition and has two more friends who do not even recognize each other, Fig.3 depicts the graph of the same scenario is 1-neighbourhood graph [6]. The vertex depicting Ed could be recognized individually in the network as no other vertices have a similar 1-neighbourhood graph.

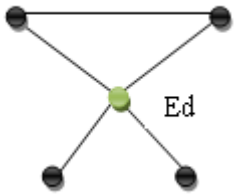

Fig.3 Ed 1-neighborhood graph

In the same manner, Ada could be recognized in Fig.2 when the adversary recognizes 1-neighborhood Ada graph. In this instance, by recognizing Ed and Ada, an adversary could identify from the released social network as shown in Fig.2 that Ed and Ada are friends and has a mutual friend. Further information as privacy could be taken in a way that how better a victim is linked with the whole network with the concerned victim position to the centre of the network. For privacy preservation, it should be noted that the individual could not be recognized as accurately in an anonymized network having a probability more than 1k (users mentioned parameter with the similar spirit in the k-anonymized model). With the addition of noise edge associating Alex and Bel, the 1-neighbourhood graph of each vertex as shown in Fig.4 is not imitable. Adversary having 1-neighbourhood knowledge cannot recognize some individual from the anonymous graph having confidence more than half [7].

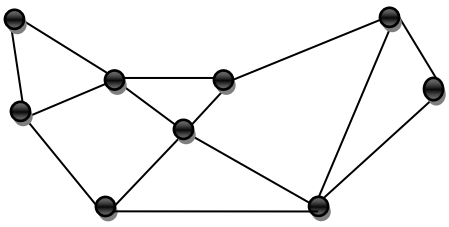

Fig.4 2-anonymous network

If the social network is improperly posted to the public, it may reveal privacy. A methodical approach is needed to anonymize social network data before it is revealed. Though, anonymizing social network data is more challenging than anonymizing relational data because of the below issues [8]:

- The initial one is that modelling of adversaries' background knowledge for social network data is more exigent as compared to relational data. In relational data, quasi-identifiers are formed by attributes set. The adversaries may utilize the quasi-identifiers for the recognization of individuals from numerous tables. Though, in the case of social networking data, it becomes more complex and tough. It is because, a small piece of information might be utilized for the identification of individuals like edges and vertices labels, induced subgraphs, neighbourhood graph or amalgamation of both.
- The subsequent one is the computation of information loss in social network anonymization is challenging as compared to relational data. It could be computed tuple by the tuple.
- The last issue is concerned in devising the anonymization techniques being a crucial task than relational data. Group anonymization for relational data of tuples does not influence other tuples within the tuples in the table. Therefore, divide and conquer techniques are utilized for anonymization of relational data. Though divide and conquer techniques might not be valuable for social network data as variation in labels of edges and vertices might not affect the other vertices neighbourhood and add or removing the edges and vertices might influence another edges and vertices with network properties. So, definite methods have to consider for the anonymization of social networking data.

Numerous privacy preservation methods are there, like K-anonymity, L-diversity, t-closeness and so on for data anonymization. Among all, K-anonymity is considered as one of the vital anonymization methods because till date, there is no re-assessment of the definite re-identification possibility of k-anonymized data [9]. This research has dealt with preserving the data using a k-anonymization method using Cuckoo Search (CS) and Feed Forward Back Propagation Neural Network (FFBPNN) to reduce the node miss placement in groups and to cross-check the structure. QoS measures are considered and are compared with the conventional ones to compute the effectiveness of the proposed work model [10-11].

B. K-Anonymity in social networking

K-anonymity [12] is one of the techniques that protect against the linkages and the recognition of the records. Each distinct tuple in the projection over quasi-identifier attributes occurs at least k −times in the k-anonymous table with the use of generalizations and the suppressions. It has two properties:

- An individual can only be linked to a group of at least k private entities in the anonymous dataset.
- An anonymous dataset represents a unique tuple in the private dataset and it doesn't have false and noisy information.

Let us suppose that a challenger recognizes the p-neighbourhood network structure of an aimed victim as background knowledge and needs to identify the vertex of the target victim in L. So, a privacy model of the k-anonymity of the social network is employed to spoil the identity attack.

Retrieval Number: J87920881019/19©BEIESP
DOI: 10.35940/ijitee.J8792.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The main idea is to confirm that the l-neighbourhood network structure of the vertex in a social network L is as same as the l −neighbourhood networks structure of minimum k-1 other vertices in L [13-14].

The definition I: Let L be a social network and k be a privacy threshold specified by social network data holder. Vertex u in L is k-anonymous only if there is a k-1 another vertex $v_1, v_2, v_3, \ldots\ldots v_{k-1} \in W$ so that M^1u and $M^1v_1, , \ldots . M^1v_{k-1}$ are isomorphic. A social network can be k-anonymous when every vertex $w\, u \in W$ in L is k-anonymous [15]. The protection provided by k-anonymity is easy and simple to understand like if any table has k-anonymity for some value m than the one who knows only the quasi-identifier of someone cannot recognize the record equivalent to the individual having confidence more than $1/k$. It provides protection against identity disclosure; it doesn't provide enough protection adjacent to attribute disclosure [16].

A. Clustering Based Anonymization

The main objective in liberating the anonymize database is to infer techniques for the prediction of private data from the public data [17]. This article has defined the clustering algorithm for K-an onymization. It initiates with the arbitrary records partitioning in clusters, later, it considers the n records in iterated manner. The records have been checked while moving from the existing cluster and while enhancing the stimulated an onymization utility. The loop iterates when it attains the local enhancement. Because there is no assurance that the process discovers the comprehensive enhancement, so, it might be iterated number of times with varied arbitrary partitions as the initial point to discover the best local enhancement between that iterated searches. The data of social networks has initiated to be examined from a definite privacy viewpoint that deemed the attribute value that distinguishes the network individual entities, the relationships with the different entities [18].

Social networking privacy preservation without disclosing the sensitive information of users is an imperative issue. Clustering based technique clusters the edges and the vertices in groups and then anonymised the subgraph in super-vertex [19]. Accordingly, the details of the individuals could be hidden appropriately.

This paper is organized in such a manner that section-II describes the literature review of the previous research work, section –III represent the problem formulation, section –IV illustrates the proposed mechanism in form of improved –anonymity and the results obtained and the comparative analysis are shown in section V and the research work crux has been defined in section VI following the references.

II. Literature Survey

The research in social networks privacy is very recent and many questions are still to be answered. Only a few researchers have explored this integrative field of privacy in social networks from a computing perspective. In this section, a short overview of the approaches is presented.

Aanchal Sharma and Sudhir Pathak (2018) [19] have utilized k-anonymity for the privacy preservation of sensitive information in social networking. The clustering concept has been used by combining the same degree in one group and the procedure is iterated till there is an identification of noisy node. For computing, the efficacies of the proposed method, parameters, such as APL and information loss are considered and a decline of information loss with 0.43% is noted. The research has not included any of the classification technique that could be used for the classification process. That is why; the information loss and APL are more. Dan Yin et al. (2017) [14] have presented a novel anonymity method termed as K-couplet anonymity. The social network data set fulfils the k-anonymity when for some node pairs; there is at least k-1 couplet with similar attributes. Two heuristics are designed and implemented for the promotion of k-couplet anonymity. The results have shown that with numerous datasets, the utility and privacy of the social network dataset are preserved with the incorporation of novel k-couplet anonymization with the association of heuristic algorithms. The effective usage of classification and optimization algorithm should be there that could produce less APL and more efficiency. Suguo Du et al. (2018) [18] has modelled privacy leakages in huge scale social network by means of economic and technical manner. With the technical manner, Markov Chain model has been used and a dynamic attack defence tree based prototype has been developed. By considering the economic manner, the static game theory has been considered for examining the definite approaches by the defender and attacker for the enhancement of utilities by means of defender/attacker cost. For the validation of the proposed work, experimentation has been done to evaluate the three real-world datasets by conducting a review of 300 volunteers that depicts the privacy risk management of existing social network social providers. Privacy protection of personal awareness could be improved as an efficient manner for lessening system risk. Wei Feng et al. (2017) [16] has presented an anonymous authentication method on the basis of group signature for the authentication at trust levels than node identities for preventing the privacy leakage and for assuring the safe Pervasive Social Networking (PSN) communication. The proposed method has achieved safe anonymous authentication by conditional traceability and anonymity by considering trusted authority (TA). The researcher has offered a system for assuring the communication between the nodes when there is no availability of TA for a few nodes. The batch signature verification has enhanced the authenticity verification efficiency on huge message amount. The assessment has shown that presented work is efficient by means of computation complexity, reliability, flexibility, privacy preservation, scalability and communication cost. The research lacks in utilizing optimization methods with some anonymization technique that might leads to more optimized results. Benjamin C. M. Fung et al. (2013) [10] proposed a technique to k- anonymize a social network dataset with the objective of preserving common sharing patterns, which is the significant kind of knowledge necessary for marketing and consumer behaviour analysis. They have conducted the experiment on three real-life datasets that areGnutella05[1], Gnutella08[2]andAdult[3].

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

For calculating the data efficacy on numerous patterns, they have measured the transformation of the common s patterns before and after the anonymization. A+ tool called MAFIA is also used to take out the frequent s patterns. The proposed technique has preserved numerous item sets for the anonymization process but should consider an appropriate aspect for the evaluation of information utility. MingxuanYuan and Lei Chen (2013) [11] have p,roposed a k-degree-l-diversity model for privacy-preserving social network data publishing. The authors have implemented distinct $l-$diversity and recursive (c, l) diversity. The KD3D model has been tested with the CORA and DBLP data sets and the KD5D has been tested on ARNET dataset. The wide experimental results show that the noise node counting algorithms are able to attain an improved outcome than the previous work by edge editing only. The protocols have to be designed for helping the publishers for publishing a united data for privacy security. Nikita Sinha and B. Annappa (2016) [21] has utilized the Cuckoo Search Diffusion model (CSDM) as a proposed method which is considered as a meta-heuristic technique as CS algorithm. This algorithm utilizes fewer measures as contrasted to conventional methods. So, the concept of parameter tuning has come out as an easy job for this algorithm and is considered beneficial for the CS algorithm. The Cuckoo search optimization diffusion architecture should be relied on discrete timestamp. Pawan R. Bhaladhare et al. (2016) [23] has presented two methods for reducing the disclosure risk and for the preservation of privacy with the utilization of clustering algorithm. An unequal amalgamation of sensitive attribute and quasi-identifier has been created initially and then equal permutation of sensitive attribute and quasi-identifier has been developed. The evaluation has been done by means of execution time and for less information loss.

As the confidential and sensitive user's information is contained by the social network data. Accordingly, sharing of information even in an indefinite manner can violate the individual's privacy. The privacy violation occurs when the confidential and private information is being revealed to the adversary. Therefore, privacy preservation when publishing user's data without disclosing the sensitive information is a significant task and hence, the concept of k-anonymization is considered in this research for privacy preservation using CS and NN.

III. PROBLEM DESCRIPTION AND PROPOSED SOLUTION

Definition 1: Get G (N, E, and P) is a graph in which there are N no of nodes connected with E edges and has a P number of sensitive labels. The problem is to equalize the degree of each node considering that the total number of added noisy nodes remains low and the diversity of the nodes remains static. It has also to be considered that data does not get too squeeze that it loses relevance.

A. Improved K-anonymity

The algorithm starts by finding the friends of the root nodes that follow identifying of the clusters, clusters in clusters and addition and deletion of nodes in the data set till $(n - 1)$ elements don't have the same degree.

Algorithm 1: Finds friends of the root node

This method aims to find the connecting node of every single unique node in the list. In all the three datasets the first column represents the main node whereas the third set represents the connecting node. It first finds all the master nodes i.e. unique nodes in the dataset and for each master node, it looks to whom it has been connected. The example for the same is shown in Table 1. and Fig.5.

Table 1. The sample dataset

1	5644	3
1	5478	4
2	1789	5
2	1788	6
2	1786	3

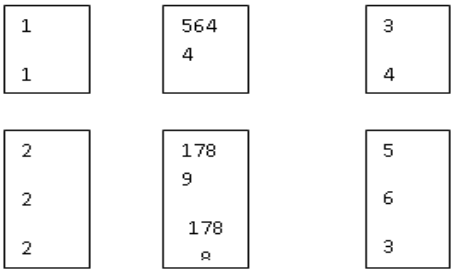

Fig.5 Friend list created

As shown in table 1. , there are two master nodes namely 1 and 2 and they have friend nodes (3, 4) and (5, 6, 3) respectively. It finds the degree of all the unique nodes of the dataset. Fig.5 presents the created friend list out of the network structure.

Function Node-friend ( )
Read data // upload by user
Master_Nodes= Data (:1); // First row of each column
***For** each m in Master_node*
previous_value=First Master_node
Previous_Node= Master_Node(n) ;
L=1;
***If** previous_value=previous_value+1*
Node_alteration=1
Node_connection [previous_value]=Uploaded_data
End
End for
Return:
1 3 4
2 5 6 3

There are three stages after algorithm 1 which subsequently creates three clusters termed as the outer clusters. The average degree of all the nodes (Degree has been calculated using Algorithm 1) would be computed first. The first cluster will contain all those elements which have a degree greater than the average degree; Cluster 2 contains the values which are less than that of the average degree and the Cluster. Cluster 3 contains those nodes which have degree exactly equal to the average degree. The third cluster might remain empty as there could be no node which has an exactly average degree. As for example, consider the following table 2:

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 2. Relationship indicator

1	5644	3
1	5478	4
2	1789	5
2	1788	6
2	1786	3

Node 1 has a degree 2 whereas Node 2 has degree 3. The average degree is $(\frac{2+3}{n}.)$ Here $n = 2$ as there are only 2 unique nodes in the master set (1, 2). The average degree is 2.5. Hence node 1 will fall in cluster 2 and node 2 will fall in Cluster 2 whereas cluster 3 would remain empty. The algorithm written below explains the prediction of the example set solved in the above-written statements. The Node Friend function is followed by the partitioning mechanism based on the average degree of the node.

Function Cluster formation ( )

1. $Function\ Create_{cluster}(data_{files})$
2. $Foreach\ dataf\ in\ data_{files}$
3. $Degree_{InOut} = Identify_{degree}$
4. $Average_{Degree} = \sum_{j=1}^{m} \frac{deg}{m}$
5. $Create_{two}group(G1, G2);$
6. $G1 > Degree_{InOut}$
7. $G2 < Degree_{InOut}$

Once the degree of the entire data is calculated, two separate groups are created, one having less than the average degree and other having higher than the average degree. Putting the nodes in the separate group reduces the search space but does not ensure the exact anonymisation in the network. Here Cuckoo Search Algorithm is applied to anonymize the grouped elements.

Algorithm 2: Cuckoo Search

1. $Application_{Cuckoo}\ (Group_{Elements})$
2. $Input: Group_{Elements}, Output: anonymized$
3. $Current_{Egg} = Key_{parents}(Group_{Elements})$
4. $Other_{Eggs} = Children(Current_{Egg})$
5. $If\ Current_{Egg}.Connection_{Count} < Group_{Elements}.Average_{Degree}$
6. $Add_{Connection_{Cukoo_{Fitness}}} = F_x(Current_{Egg}, Other_{Egg}, 1)\ //\ flag = 1$
7. $Elseif\ Current_{Egg}.Connection_{Count} > Group_{Elements}.Average_{Degree}$
8. $Remove_{Connection_{Cukoo_{Fitness}}} = F_x(Current_{Egg}, Other_{Egg}, 2)\ //\ flag = 2$
9. $End\ If$

$Where\ F_x is\ the\ fitness\ function\ of\ Cuckoo_{Search}$

Table 3. Fitness function of Cuckoo Search

Fx Output	$Fitness_{Terms}$
Connections to Add	$Flag = 1\ Intersect_{Common}(Other_{Egg} > G_{Group_{Elements}}, Current_{Egg}$
Connections to Remove	$Flag = 2\ Intersect_{Common}(Other_{Egg} < G_{Group_{Elements}}, Current_{Egg}$

The Cuckoo Fitness works in two stages as shown in Table 3. The first stage is when the $\text{Current}_{\text{Egg}}$ (the element of the group) requires to add connection and the second stage is when the $\text{Current}_{\text{Egg}}$ wants to remove the connection. Flag Value 1 denotes that the node wants to increase the degree and Flag value 2 denotes that the node wants to decrease the degree. In the same cluster, if the nodes have similar group connections and other has to decrease its degree then the mutual connection will be removed from the parent and will be added to the child or demanding egg. In a similar fashion, the second value is when the parent wants to lose. In this scenario, the common intersected eggs will be added to the child and will be removed from the parent.

In addition to Cuckoo Search, the outcome of the Cuckoo Search is cross-validated utilizing Feed Forward Back Propagation Neural Network. The instances of Neural Networks are shown in Table 4.

Table 4. Neural Architecture

Total Neuron Count	50
Propagation Model	Levenberg
Propagation Type	Feed Forward
Cross validation Parameter	Mean Square Error (MSE)

Algorithm 3: Feed Forward Back Propagation Neural Network

Initialize ANN with parameters *– Epochs (E)*
– Neurons (N)
– Performance parameters: MSE, Gradient, Mutation and Validation Points
– Training Techniques: Levenberg Marquardt (Trainlm)
– Data Division: Random
For *each set of Training Data // Cuckoo search returned data*
Group = Categories of Training data
End
Initialized the FFBPNN using Training data and Group
Net = Newff (T, G, N)
Set the training parameters according to the requirements and train the system
Net = Train (Net, Training data, Group)
Classify = simulate (Net, test properties)

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Return: *Output to validate the cuckoo search*
End

Algorithm 3 is used to validate the cuckoo search algorithm based on training mechanism and the used parameters of FFBPNN are given in the Table 5. Using the FFBPNN, the performance of proposed system became better as compared to the existing work which is well defined in the next chapter.

IV. RESULTS AND DISCUSSION

This section has been categorized in two segments. In the initial segment, the outcome of the proposed work has been delineated and the subsequent section has shown the comparison of proposed work with the conventional techniques. For the comparison, the research work of [19] and [23] has been considered that has utilized APL and Information loss as the computing measures.

A. Result analysis of proposed work

This segment explains the results obtained after the evaluation of the proposed work. For the analysis, measures, such as APL and Information loss has been computed. The mathematical expression and description of the considered parameter are defined beneath:

a) Average Path Length (APL)

The average path length is the ratio of the distance between two nodes with diverse information to a total number of nodes in the dataset.

$$APL_{G(L_1 \text{ and } L_2)} = \frac{\sum_{\forall n_{i.s=L_1}, n_{j.s=L_2}} d(n_i, n_j)}{\sum_{\forall n_{i.}=l_{i,}, n_{j.}s=l_{2,.}} 1}$$

As shown in equation (1), N is the total number of nodes and n is the current node.

b) Information loss

Involvement in social networking results in personal information loss. Better encryption directs to information loss in less amount.
Below, the results are shown that are computed after the simulation work.

Table 5.APL evaluation of proposed work

Number of nodes	APL
1	0.11
2	0.19
3	0.32
4	0.30
5	0.33
6	0.31
7	0.29
8	0.30
9	0.24
10	0.30

Fig.6 and Table 5. shows the results of the proposed work for APL. The accurate calculation of APL helps in analyzing the shortest path among random nodes. Though, attaining intact social network is complex as of privacy and security in restrictions of privacy protection.
The x-axis in the figure shows the k value of nodes from 1 to 10 whereas Y-axis illustrates the obtained values of APL. APL is defined as the measure for computing the effectiveness of the information or the huge transmission of data on the social network. From the graph, it has been seen that the APL for the proposed work is less with an average of 0.269.

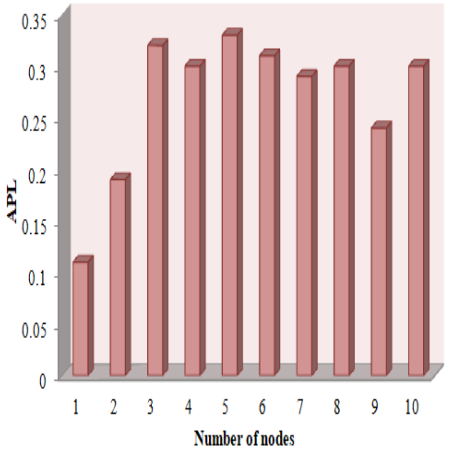

Fig.6 APL for proposed work

B. Comparative analysis of proposed work with the conventional techniques [19] and [23]

This section elaborates the comparison of proposed mechanism with the existing mechanism to depict the effectiveness of the work. For the comparison, [19] and [23] has been considered.

Table 6.Information loss evaluation for proposed work

Number of nodes	Information Loss
1	30.5
2	29.6
3	30.3
4	31.5
5	29.3
6	30.4
7	31.7
8	30.4
9	32.2

Fig.7 and Table 6. shows the results of the proposed work for Information Loss. Less information loss leads to more accurate decision making. The information loss in communication network occur when more links or nodes in physical communication network fades away.

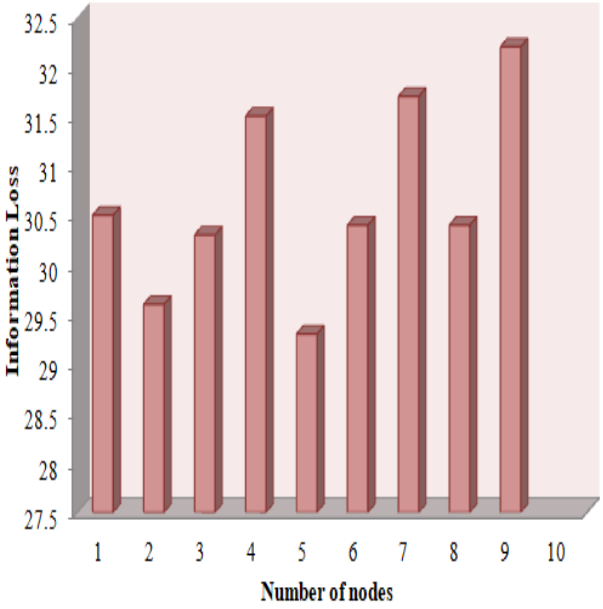

Fig.7 Information Loss for proposed work

The x-axis in the figure shows the k value of nodes from 1 to 9 whereas Y-axis illustrates the obtained Information loss values.

Table 7.Comparison of APL for proposed and existing work [19]

Table 7.Comparison of APL for proposed and existing work [19]	
Table 7.Comparison of APL for proposed and existing work [19]	**Existing [19]**
Table 7.Comparison of APL for proposed and existing work [19]	0.15
Table 7.Comparison of APL for proposed and existing work [19]	0.22
Table 7.Comparison of APL for proposed and existing work [19]	0.36
Table 7.Comparison of APL for proposed and existing work [19]	0.35
Table 7.Comparison of APL for proposed and existing work [19]	0.37
Table 7.Comparison of APL for proposed and existing work [19]	0.36
Table 7.Comparison of APL for proposed and existing work [19]	0.34
Table 7.Comparison of APL for proposed and existing work [19]	0.31
Table 7.Comparison of APL for proposed and existing work [19]	0.28
Table 7.Comparison of APL for proposed and existing work [19]	0.41

More information loss leads to less reliability of the novel mechanism, so, it should be less for more privacy of the system. The average value of information loss for the proposed work is 30.65 which are quite productive by means of its effectiveness.

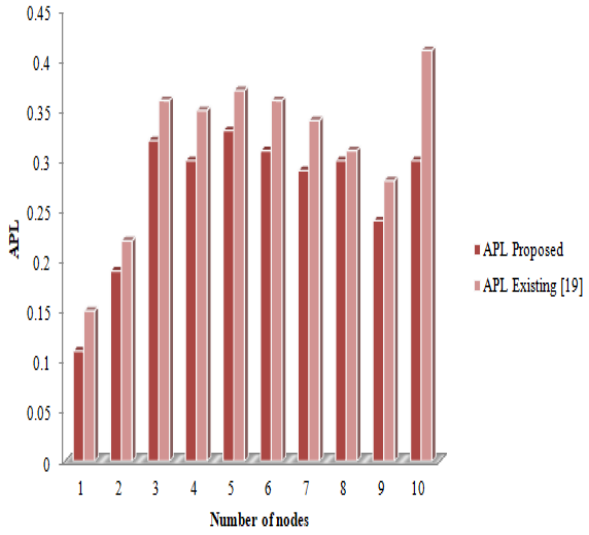

Fig.8 Comparison of APL

The comparison of APL for proposed and existing work [19] for k=10 has been delineated in Fig.8 and is shown in tabular form in table 7. There is a reduction of 14.6% in APL of the proposed work as contrasted to existing work. Therefore, it is evident from the graph that proposed work has outperformed in terms of APL than conventional technique.

Table 8.Comparison of Information Loss for proposed and existing work [19] and [23]

Information Loss		
Proposed	**Existing [19]**	**Existing [23]**
30.5	32	33
29.6	33.1	33.3
30.3	33	33.5
31.5	33.2	33.6
29.3	33.8	34
30.4	33.1	34.2
31.7	34.6	35
30.4	34.1	35.2
32.2	35.8	36

The comparison of information loss for k as 9 ha been shown in Fig.9 and the values has been depicted in table 8. The comparison has been contrasted for proposed and existing work [19] and [23]. It can be seen from Fig.9 that the information loss of the proposed work is less as compared to the conventional methods [19] and [23].

Retrieval Number: J87920881019/19©BEIESP
DOI: 10.35940/ijitee.J8792.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

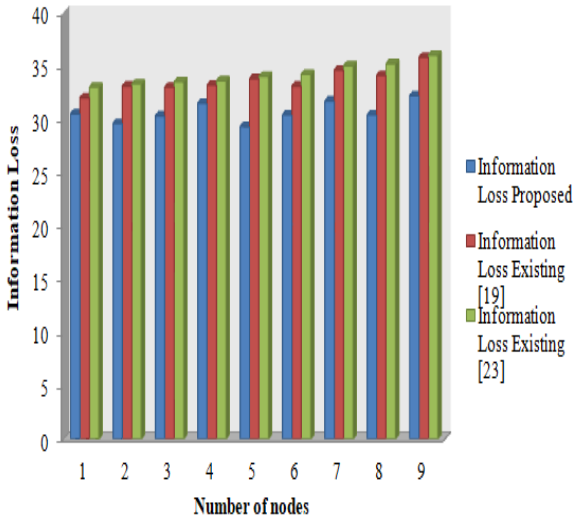

Fig.9 Comparison of information loss

The average value of information loss for [19] is 33.63 and for [23], it is 34.2. 8.61% of reduction has been noted in information loss of proposed work with [19] and 10.38% of reduction has been seen while contrast with [23]. From the assessment, we came to the conclusion that information loss in case of proposed work is very much less than the existing techniques.

V. CONCLUSION

This article has provided a novel mechanism of privacy preservation in social networking. An enhanced k-anonymity method has been proposed by considering the advanced clustering principle. Cuckoo search optimization method has been used that has designed a novel fitness function. The outcome of the Cuckoo Search has been cross validated by means of Feed Forward Back propagation neural network. Improved K-anonymity has been considered for finding the friends of root nodes for the cluster identification and for connecting the node of each single node. The role of cuckoo search algorithm is to group the element anonymization. QoS parameters, viz. APL and information loss are considered to depict the outcome and the comparison has been drawn with the conventional techniques. The comparison of APL has been contrasted with [19] and [19] with [23] are considered for the comparison of Information loss to illustrate the effectiveness. There is a reduction of 14.6% in APL of the proposed work whereas 8.61% of decline has been noted with [19] and 10.38% of fall has been noted in information loss.

REFERENCES

1. A. Kaur, “A hybrid approach of privacy preserving data mining using suppression and perturbation techniques”, In International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, pp. 306-311,2017.
2. D. Patel and R. Kotecha, “Privacy Preserving Data Mining: A Parametric Analysis”, In Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications, Advance in Intelligent Systems and Computing, Vol. 516, pp. 139-149, 2017.
3. P. Mohana Chelvan and K. Perumal, “Stable Feature Selection with Privacy Preserving Data Mining Algorithm”, Advanced Informatics for Computing Research. Communications in Computer and Information Science, Springer. Singapore, Vol. 712, pp 227-237, 2017.
4. Y. Song, P. Karras, Q. Xiao and S. Bressan, “Sensitive Label Privacy Protection on Social Network Data”, IEEE transactions on knowledge and data engineering, Vol.25, No.3, pp 562-571, 2013.
5. Zhou and J. Pei, “The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks”, Knowledge and Information Systems, Vol.28, No.1, pp 47–77, 2010.
6. K. Ilavarasi and B. Sathiyabhama, “An evolutionary feature set decomposition based anonymization for classification workloads: Privacy Preserving Data Mining”, Cluster Computing, Vol. 20, No. 4, pp 3515–3525, 2017.
7. G. Priyanka, P. Darshana and Kotecha Radhika, “Privacy-Preserving Associative Classification”, In International Conference on Information and Communication Technology for Intelligent Systems. Smart Innovation, Systems and Technologies, Springer, Cham, Vol. 2, pp.245-251, 2017.
8. X. Wu, X.Ying, K. Liu and L. Chen, “A survey of privacy-preservation of graphs and social networks”, In Managing and mining graph data, Springer, Boston, MA, pp. 421-453, 2010.
9. K. LeFevre, D. J. DeWitt and R. Ramakrishnan, “Mondrian Multidimensional K Anonymity”, In IEEE International Conference of Data Engineering, Vol. 25, pp.1-11,2006.
10. B. C. M. Fung, Y. Jin and J. Li, “Preserving privacy and frequent sharing patterns for social network data publishing”. *In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, Niagara Falls, ON, pp. 479-485, 2013.
11. Mingxuan Yuan and Lei Chen, “Protecting Sensitive Labels in Social Network Data Anonymization”, IEEE transactions on knowledge and data engineering, Vol. 25, No. 3, 2013.
12. Narayanan and V. Shmatikov, “De-Anonymizing Social Networks”, Proc. IEEE 30th Symp. Security and Privacy, pp. 173-187, 2009.
13. Z. He, Z. Cai and J. Yu, “Latent-Data Privacy Preserving With Customized Data Utility for Social Network Data”, In *IEEE Transactions on Vehicular Technology*, Vol. 67, No. 1, pp. 665-673, Jan. 2018.
14. D. Yin, Y. Shen and C. Liu, “Attribute Couplet Attacks and Privacy Preservation in Social Networks”, in *IEEE Access*, Vol. 5, pp. 25295-25305, 2017.
15. Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin and K. Ren, “Real-Time and Spatio-Temporal Crowd-Sourced Social Network Data Publishing with Differential Privacy”, In *IEEE Transactions on Dependable and Secure Computing*, Vol. 15, No. 4, pp. 591-606, 2018.
16. W. Feng, Z. Yan and H. Xie, “Anonymous Authentication on Trust in Pervasive Social Networking Based on Group Signature”, In *IEEE Access*, Vol. 5, pp. 6236-6246, 2017.
17. M. Siddula, L. Li and Y. Li, “An Empirical Study on the Privacy Preservation of Online Social Networks”, In *IEEE Access*, Vol. 6, pp. 19912-19922, 2018.
18. S. Du , X. Li , J. Zhong , L. Zhou , M. Xue and H. Zhu, “Modeling Privacy Leakage Risks in Large-Scale Social Networks”, in *IEEE Access*, Vol. 6, pp. 17653-17665, 2018.
19. Aanchal Sharma and Sudhir Pathak, “Enhancement of k-anonymity algorithm for privacy preservation in social media”, International Journal of Engineering & Technology, Vol. 7, No. 2.27, pp.40-45, 2018.
20. Amir Hossein Gandomi, Xin-She Yang and Amir Hossein Alavi, “Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems”, Engineering with Computers, Vol. 29, No.1, pp 17–3, 2013.
21. Nikita Sinha and B. Annappa, “Cuckoo Search for Influence Maximization in Social Networks”, In Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics, Springer, New Delhi, Vol 44. pp 51-61, 2016.
22. Weibo Liu,Zidong Wanga, Xiaohui Liu, Nianyin Zeng, Yurong Liu andFuad E.Alsaadi, “A survey of deep neural network architectures and their applications”, Neurocomputing, Vol. 234, No.19, pp.11-26,2017.
23. P. R. Bhaladhare and D. C. Jinwala, “Novel Approaches for Privacy Preserving Data Mining in k-Anonymity Model”, Journal of Information Science and Engineering”, Vol. 32, No.1, pp. 63-78, 2016.

Retrieval Number: J87920881019/19©BEIESP
DOI: 10.35940/ijitee.J8792.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)