Enhancing Privacy Preservation Using Hybrid Approach Of K-anonymity, Artificial Bee Colony And Neural Network

Shivani Sharma[1], Sachin Ahuja[2]

[1,2]**Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab, India**

Abstract: The rising popularity of social networks has also raised the risk adjoining the dissemination of the user's personal information over the network. This has raised the demand of privacy protection. Privacy preservation is the rising issue in the social networks that are the hot spots where information theft instances are very common. The present approach focuses the protection of sensitive information based on k-anonymity. K-anonymity is one of the most popular approaches privileged by graphs and nodes functionality. The proposed study is based on the enhancement of k-anonymity by focusing at node level to address the privacy protection issue. The process involves first identification of sensitive nodes and then applying optimization techniques. At this step, authors have introduced Neural Network (NN) and Artificial Bee Colony (ABC) in order to reduce the node miss placement in the groups. The study is evaluated against k-anonymity on small and larger datasets in terms of average path length and information loss. Comparative analyses have shown APL reduction of 1.636 and 1.371 is achieved using ARNET and SDFB datasets over 900 nodes. Additionally, optimization also resulted in average information loss reduction of 0.57% and 8.95% was observed for small and larger datasets.

Index Terms: Privacy preservation, K-anonymity, Neural Network, ABC, APL, Information Loss.

1 INTRODUCTION

Internet is playing a indispensable role in the popularity of social networking. Numerous big data resources, especially social networking sites regularly share huge amount of data over the network. This data exhibits front and backend characteristics and is basically disseminated with the sole purpose of analysis [1]. Worldwide analysis results of com Score shows that U.S. people use 98% of the available time for surfing over Instagram [2]. The time has shown a tremendous amount of increase in both scalability and variety of data over network. The statistical analyses have shown that network sites such as twitter cover 600 million users with 0.5 billion users actively tweets. Similarly, Facebook exhibits at least 1.65 billion users out of which 1 billion accounts to the most active users. When discussion the popularity of Amazon, it is found that it has 304 million users who deal with 9.65 billion items over the year. Tencent QQ has a comparatively low popularity and accounts to 829 million active users. This shares huge amount of data though social networking channels and websites. [3].The stats shows the growing strength of Social media in connecting people over the globe. Example of social media is shown in Figure 1.

***Fig. 1**. Social media over globe.*

This big data forms a rich source of information and is very advantageous for conducting various type of analysis. On the other hand, the rising numbers of users over the network have also raised the privacy risk and incidents of various types of theft and attacks. Hence, the social networks have been the major victims[4, 5]. The users over these networking sites share the information under various attributes like gender, location, contact information, etc. This personal information can get compromised due to malicious act that severely violates the integrity of the data and privacy protection policy [6]. As a result, it has become mandatory for a service provider to offer privacy protection before publishing any kind of data over the network. Data comprising sensitive personal information have been the mainly focused. The observed inconsistency among the data instances has raised the foes that keep their eye over the sensitive information that is integrated in the published records [7].

- *Shivani Sharma is Ph.D. student at Chitkara University. Her research interests include Privacy preserving in Data Mining, K-anonymity, ABC and Neural Network. She has done graduation in Computer Science from Punjab University and postgraduation from Chitkara University.*
- *Sachin Ahuja is Director Research at Chitkara University, Punjab. Presently working in the field of Data Mining, Artificial Intelligence & Machine Learning. He has done his Master in Computer Science & Engineering and Doctorate in Data Mining.*

The paper is organized in six sections including introduction. In Section 2, problem formulation and key motivation behind the study is discussed, section 3 summarized the related work and in section 4 proposed design is discussed. Section 5 covers the results and discussion. The paper is concluded in section 6 while discussing the core finding.

2 PROBLEM FORMULATION AND KEY MOTIVATION

The authors are inspired by the social network analysis (SNA). In the process individuals are represented by the nodes and the network connection corresponds to individual communication. The adversaries took its advantage while understanding of the individual position in terms of vertex of a social networking platform. For instance, a social network of 6 individuals is shown in Figure 1. Here, each node represents an individual in the social network [8]. Individuals are depicted as A, B, C, D, E and F.

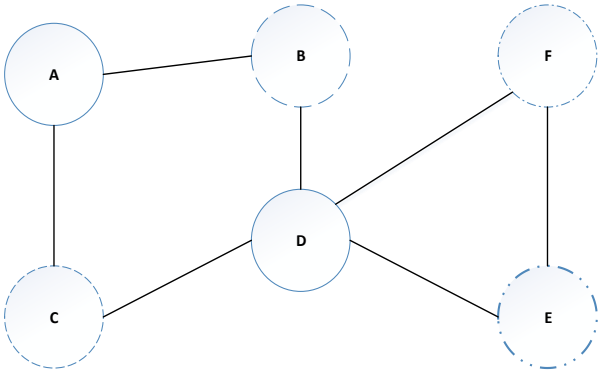

Fig.2*. Social Network*

Figure shows the inherit individuality of five friends in the social network. In order to achieve privacy protection, the very first step to hide the individual identity over the network. This results in the formation of an anonymous network as illustrated in Figure 2.

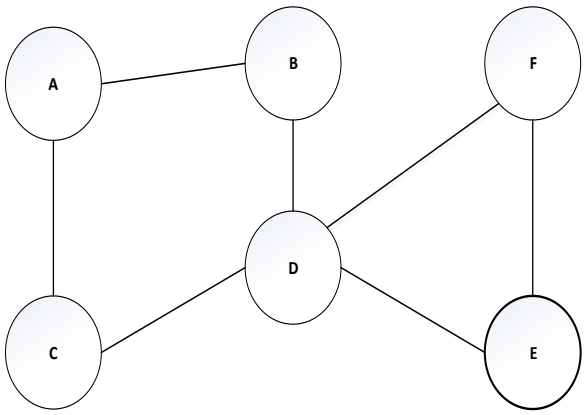

Fig.3*. Anonymous social network*

In the anonymous network too, the knowledge of individual's neighborhood could result in the leakage of one's personal information to the foe. For instance, let's consider a social network of individual 'D' with two mutual friends 'E' and 'F' and two direct friends 'B' and 'C' as shown in Figure 3. The truncated social network corresponds to 1-neighabourhood graph [9]. In this scenario, adversary could easily recognize individual 'D' as in the network no other individual exhibits a similarity in terms of D -1-neighabourhood graph.

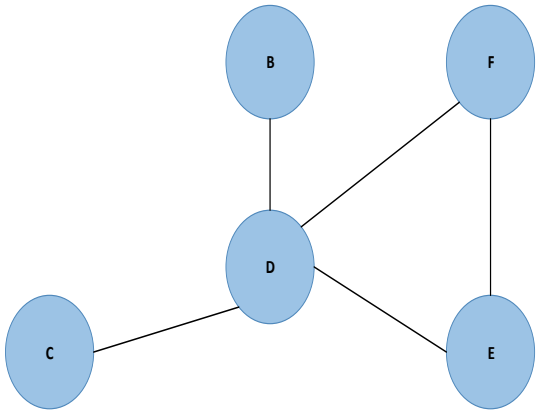

Fig.4*. D – 1-neighbourhood graph*

In the similar fashion, foe could also recognize individual 'C' in the social network shown in Figure 2 on the basis of 1-neighabourhood graph of 'C'. This means that by recognized two individuals 'C' and 'D' of the social network foe could establish that individuals 'C' and 'D' are friends. Moreover, high dimensional information could be retrieved by analyzing the position of the individual in the social network. It is observed that high privacy protection can be achieved for the individuals with high dimensionality in anonymized network as illustrated in Figure 4. Additionally, association of edge noise also adds a step in protecting the individual identity.

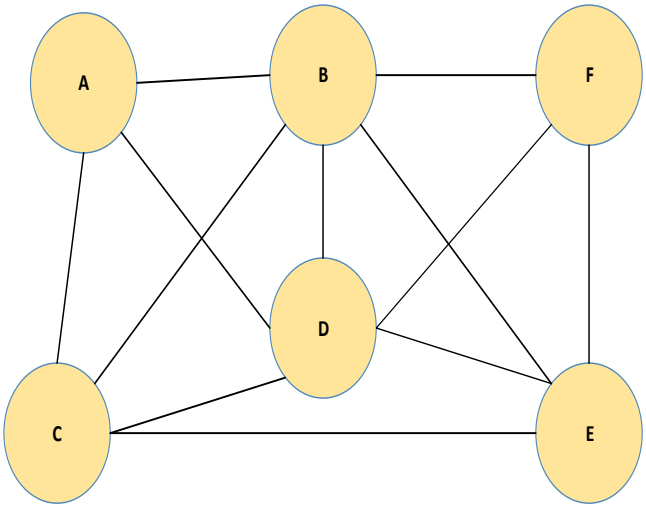

Fig.5*. Noisy anonymous network*

In a noisy anonymous network, ability of foe to correctly identify the individuals is challenged to half [10]. Privacy is the main concern of social networks that are inappropriately shared publically. Social networks require to be systematically anonymized in order to protect their privacy.

The major issues related to anonymization of social network in comparison to the anonymization of relation data [11] are as follows:

1. It is difficult to represent the background knowledge of foe for social network as compared to the relation data. In case of relation data, foe identifies the individuals by analyzing the tables on the basis of quasi-identifiers. While in case of social networks, it becomes difficult to analyses individual identity on the basis of edges and vertices, neighborhood graphs or subgraphs.
2. In case of social network, information loss is a very demanding parameter that rises with rising complexity of the network in comparison to the anonymized relation data.
3. The development and formulation of anonymization technique is a vital step in social networks than relation data. In case of group anonymization tuples are usually independent and do not control each other behavior in relation data whereas in social networks, addition and removal of edge and vertices greatly modifies the whole network properties. Hence, a well-organized and defined anonymization technique holds top position in anonymization of social network.

A number of privacy preservation techniques are available involving t-closeness, L-diversity, k-anonymity, etc. Out of these techniques there is no instance available to consider the re-evaluation of the k-anonymity based anonymized. Hence, k-anonymity holds the top position among the available techniques [12]. The present research focuses the privacy preservation of the data using k-anonymization technique using Neural Network (NN) and Artificial Bee Colony (ABC) in order to decrease node miss-placement among groups and to recheck the structure. Quality of service is measured in terms of APL and information loss is compared to evaluate the effectiveness of the proposed design. [13-14]

2.1 Anonymity in Social Networks

The technique deals with anonymization of social network data [15]. It covers the protection to the relationships while diminishing the act of tracing and identifying the individual information [16]. The information corresponds to a single attribute that subsist k-times in a k-anonymous based privacy protection data. It exhibits dual properties, firstly, individual in the social network can be linked to group that also exhibits k-entities and secondly, the anonymous dataset corresponds to noise free information attributes. For instance, let's consider that foe recognizes a social network that has p-neighborhood network structures. The foe focuses the 'M' individual based on the background knowledge and p-neighborhood information. Here the privacy preservation k-anonymity model for social network is used to offer protection against the attack. The goal of this defense mechanism is to identify if 1-neighabourhood network of 'M' vertex or individual is similar to M-neighborhood networks structures [17-18].

Definition 1: Let 'S' corresponding to a social network with 't' as privacy threshold assigned by the data holder of the social network. The vertex 'M' in the social network 'S' is considered to be k-anonymous only if there exists k-1 another vertex represented by 'n'

Where, $n_1, n_2, n_3, \ldots \ldots n_{k-1} \in S$

Such that, $M^1 u$ is isomorphic to $M^1 n_1, , \ldots . M^1 n_{k-1}$

A social network is said to be k-anonymous only when all the vertex in 'S' social network are k-anonymous [19].

K-anonymity based privacy protection can be understood in terms of tables. For example, k-anonymity of a value 'M' can only disseminates the records with a confidence level of $1/k$. It protects the identity but offers less protection in terms of attribute disclosure [20].

3 RELATED WORK

Related studies have shown that numerous approaches have been proposed for privacy protection. Recently, Zhang et al. in 2019 addressed the privacy issue with the engagement of multi levelled caching and spatial k-anonymity design. Initially next location of the query was predicted based on Markov approach. Further knowledge of location was used by authors to increase the location privacy based on another spatial k-anonymity design. Simulations analysis demonstrated a high magnitude privacy protection and shown success in terms of reduced transparency of location based server [21]. Sharma and Pathak in 2018 employed the advantage of k-anonymity to offer protection to sensitive information disseminated over the social networks. The authors had used the concept of clustering. In this approach clustering is repeated until it encounters a noisy vertex. The evaluation was done in terms of APL and information loss. The results demonstrated a 0.43% reduction in information loss. The authors added that involvement of a classification technique could improve information loss parameter [22].Wei Feng et al. in 2017 has proposed anonymous verification technique that worked on utilizing group signatures in order to deal with privacy outflow while offering Pervasive Social Networking (PSN) communication. Safe levels of endorsement were achieved with the involvement of conditional traceability and anonymity by considering trusted authority (TA) [20]. Bhaladhare et al. in 2016 designed a technique to decrease the information loss due to systematic clustering. In the process 'Greedy k-member' and 'Systematic clustering' were also discussed. The attribute information was used to create anonymized data. Privacy protection of the data was done by using systematic clustering method while disclosure risk was dealt by greedy technique. In the experimental evaluation UCI machine learning data sets comprising of 32561 records having 15 attributes were employed. The evaluation was done in terms of execution time and information loss. The results demonstrated the proposed technique resulted in lesser information loss [23].Tsai et al. in 2015 edge based approach to deal with privacy using k-anonymization in terms of finding the shortest path between the nodes. They believed that there should be a minimum of k number of shortest edges or paths between the two nodes representing a destination node and a data sensitive node. In

light of this fact and belief authors proposed three algorithms that were focussed to address three distinct edge categories. This resulted in the k-shortest path based privacy protection with variable degree of information loss and execution time. They offered that their proposed models could be used as a base reference in order to achieve shortest path based anonymization research [24]. Triparty et al. in 2014 focused their research to deal with the privacy concerns of social networks and presented GASNA as an anonymization approach for social networks. The technique was based on greedy algorithm that offers protection of attributes in terms of l-diversity and k-anonymity of the data. Authors had also addressed the issues faced by the existing approaches to deal with the privacy in social network and recommended a few possible solutions. A partial anonymity model was proposed by the authors that successfully addressed the d-neighborhood problem when d>1 [15].Chester et al. in 2013 were majorly concerned by k-anonymization approach in social networks. Their study revolved around higher node degrees with node set modifications preferred over edge set. This approach proved to be very advantageous when real-time node labelled graphs were studied. The authors established that a very little distortion was observed in the clustering coefficients as a result of anonymization [25].

4 METHODOLOGY

4.1 Proposed Design

The proposed design involves sensitive attribute selection when the data is to be sent for the optimization. The k-anonymity part is not concerned about the sensitive attribute selection as it is only based on the identity count. Hence first of all, before sending the data for the optimization, it is required to check the sensitive attribute. Flowchart shown in Figure 6, illustrates the working of the sensitive information calculation.

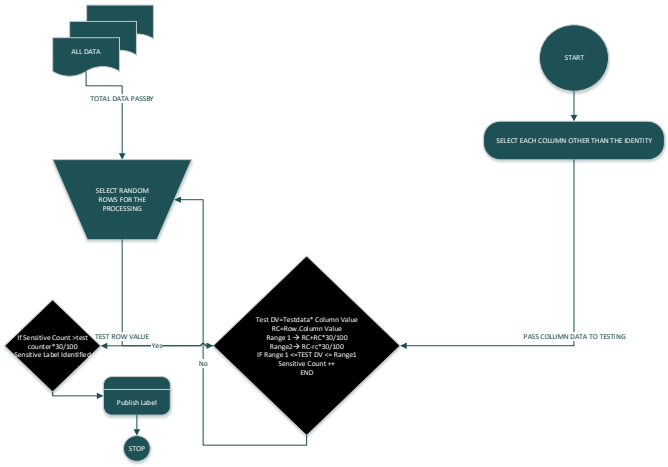

***Fig.6**. Steps to Identify Information Sensitivity*

The above flowchart works on the basis of two range formulas thatare applied in order to check whether the data is sensitive or not. Each column other than the identity column is processed and cross validated by every other label in the list. A random row count is initialized and the current value is compared with 30% margin range. If the data lies within the margin range, it is said to be possibly contain sensitive information. For instance, suppose randomly 40 rows have been selected in total and 25 times it results in possible sensitivity.The calculated sensitivity percentage for this will be $25 * \frac{100}{40} = 62.5\ \%$ which is higher than 40. Hence, this label will be considered to be sensitive.

```
function [labels] = calculatesensitive(datatocheck)
%CALCULATESENSITIVE Summary of this function goes here
%   Detailed explanation goes here
labels=[];
alltestrows=[];
sensitivecount=0;

[r,c]=size(datatocheck);
for i=1:c
   if i~=1   % id can not be sensitive attribute
      positivecounter=0;
      currentattribute=datatocheck(1,i);
      testrows=round(100*rand); % taking the random test rows
     for j=1:testrows
      alltestrows(j)=round(r*rand); % calculating the rows which will be
      if alltestrows(j)==0
         alltestrows(j)=1;
      end
     end
```


***Fig.1.** Pseudo Code for Sensitive attribute identification*

The code shown in Figure 6 corresponds to the syntax used in identifying the sensitive information using based on the flowchart described in Figure 5. Figure highlights the variables that are used for input data source, labels, sampled test data and random selection of test rows attributes in the above code. In the next step, this label information is passed to Artificial Bee Colony (ABC).

4.2 Artificial Bee Colony (ABC) algorithm

ABC algorithm is inspired by the intelligence characteristic of the honeybees. This bio-inspired algorithm exhibits admirable search ability to deal with more complicated problems.
The strength of ABC algorithm lies in the three working bees:

a) Employed Bee
b) Onlooker Bee
c) Scout Bee

The employed bee is the food collector bee. It searches the food and passes on the food to the Onlooker Bee for the checking and variations. The scout bee is mainly the resting bee and also termed as unemployed bee. It is a non working element in the proposed case.

The sensitive labels obtained in the last step are passed as the input to ABC algorithm. The flow architecture of ABC algorithm is shown in Figure 7.

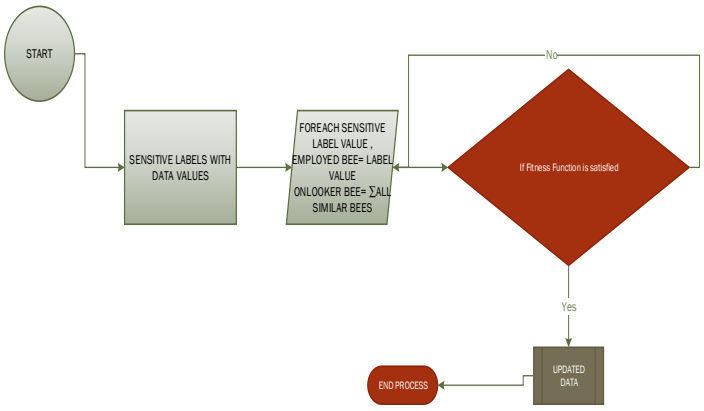

***Fig.2**. ABC algorithm flow chart.*

The fitness function of the ABC is as follows

$$f = 1 \qquad if Employed_{Bee} * Travel_{Time} < Onlooker_{Bee} * Waiting_{Time}$$

$$0 \qquad \text{Otherwise}$$

The proposed work uses Supervised Neural learning approach for the cross validation of the output of the ABC algorithm. Neural network takes the feature value and the associated label as the processing value. If the processed value matches the targeted label value only then it is considered to be in productive value.

4.3 Feed Forward Back Propagation Neural Network (FFBPNN)

Neural Network is also employed in the current methodology in addition to ABC algorithm. It is a machine learning technique that works on the basis of learning from examples or the training dataset. The results of ABC are compared with the NN. Figure 8 shows the propagation neural architecture used in the current methodology. Figure 8 shows that the Neural Network works on 20 neurons whose performance is evaluated on the basis of Mean Square Error obtained in the simulation process. The simulation progress parameters namely, performance, gradient and estimated time corresponds to 9th iteration out of a total of 1000 iteration steps.

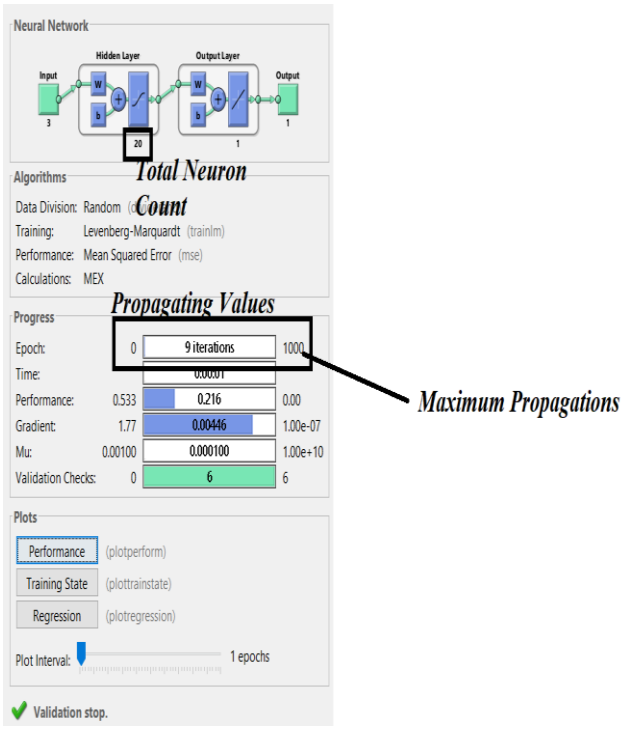

***Fig.3**. Neural Network*

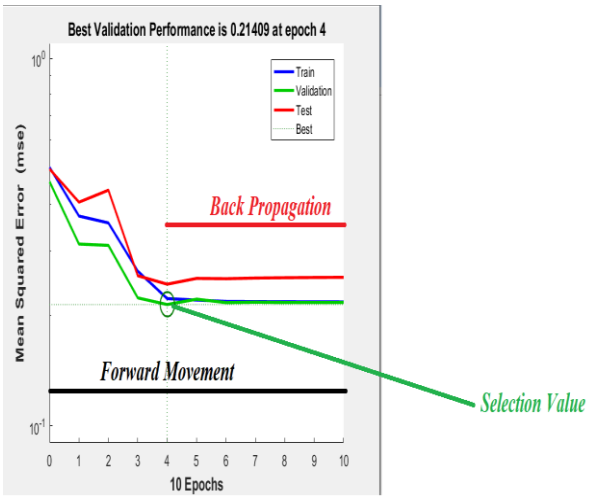

***Fig.4**. Validation Graph*

The graph shown in Figure 9 represents the validation performance of FFBPNN. The graph shows the comparative plot of training, validation and test samples over the 10 epochs verses the resultant mean square error. The best fit is observed that at epoch 4 that exhibits the performance of 0.21409.

5 RESULTS

This section is dedicated for the experimental calculation of average path length and information loss over the number of nodes in the considered social network. This is followed by the evaluation against the existing methodologies.

5.1 Parametric Calculations

The mathematical calculation of the parameters is described as follows: Average Path Length (APL) It the average value corresponding to the distance between the two vertices or the nodes path length is the ratio of the distance between two nodes to a total number of vertices or nodes present in the dataset. The average path is used to measure the connection among the two labels. Precise calculation of APL helps in analyzing the shortest path among number of random nodes. Let two levels are defined as L1 and L2. Let us assume an un-weighted graph 'G' comprises of 'A' number of vertices. Let G(a1,a2), where a1,a2ϵA indicates the smallest distance between (b/w) a1 and a2. Let us assume that G(a1,a2)=0 if a2 does not reached from a1, then the APL can be written mathematically as:

$$APL_G = \frac{1}{a \times (a-1)} \times \sum_{i \neq j} G(a_i, a_j)$$

Here, 'a' signifies the number of vertices in graph 'G'.

Information loss

Information loss in a social networking is concerned with the loss of personal information while being a part of social media network. To minimize this loss privacy protection techniques are used. The simulation results are discussed in the next section.

5.2 Evaluation of APL and Information Loss

The work involves ARNET and SDFB datasets to achieve two fold evaluations:

1. The first evaluation is done on a small data value attribute when total numbers of nodes are 10 and the kth degree is 10 at maximum.
2. The second proceeding is for the big data attribute value which contains not more than 1000 nodes.

5.2.1 APL comparison

APL is the most important parameter that help in achieving an intact social network, which is a very complex to understand in terms of privacy and security. Table 1 summarized the APL results of k-APL and APL obtained by proposed model on small ARNET data. It is observed that the proposed architecture results in lower APL as compared to k-APL over the 9 nodes. For comparative analysis the results are plotted in FIGURE 10.

TABLE 1

APL COMPARISON ON SMALL ARNET DATASET

Node Count	k-APL	APL-Proposed
1	4.350318	2.746533
2	3.662639	2.847429
3	4.882806	4.699111
4	3.097866	0.930348
5	4.526420	3.396932
6	5.099641	4.740576
7	6.033248	5.088301
8	3.842616	2.270624
9	4.728919	3.199738

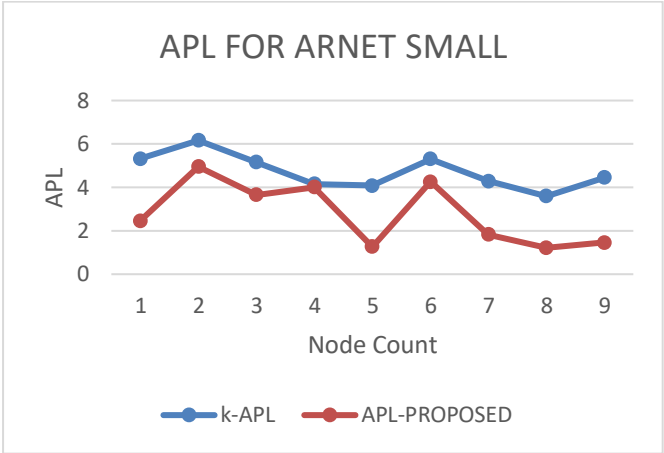

Fig. *5. APL comparison on small ARNET data.*

k-APL and APL of proposed architecture are plotted in Figure 10. Node counts ranging up to 10 node counts are plotted on X-axis and APL values are plotted on Y-axis. The average APL for k-APL is 4.47 and for APL-proposed are 3.324. It is observed that APL of the proposed architecture is lower on all the nodes as compared to the k-APL values. k-APL and APL-proposed are also evaluated using big data of ARNET database and the corresponding values are listed in Table 2. It is observed that even on large data set with nodes ranging to 900, the proposed architecture has lower APL values.

TABLE 2

APL COMPARISON ON BIG ARNET DATA.

Node Count	k-APL	APL-Proposed
100	3.304524	0.882216
200	6.602886	4.081944
300	6.018511	4.192175
400	5.44455	4.464253
500	5.718338	2.989008
600	5.13460	4.836879
700	5.205743	4.15437
800	4.324461	2.619248
900	3.42752	2.235121

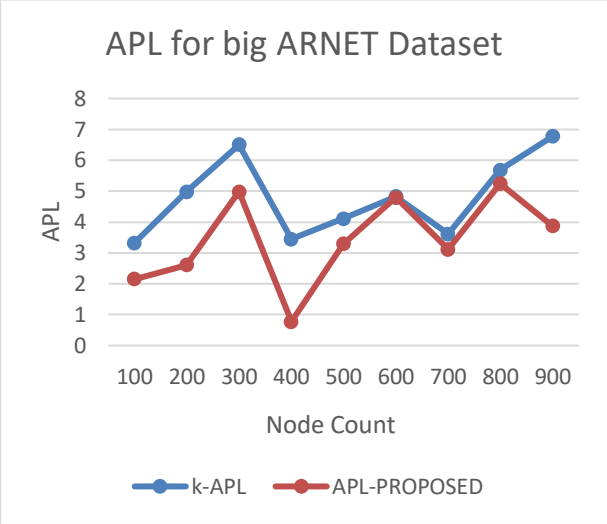

***Fig.6**. APL comparison on big ARNET dataset.*

The results of APL using big data of ARNET database are plotted in Figure 11. Node count in this case ranges from 100 to 900. The graph shows a number of fluctuations in APL values over the 900 nodes. It is observed that k-APL and APL-proposed achieves an average value of 5.02 and 3.384, respectively. It means that APL reduction of 1.636 is achieved by the proposed work as compared to the existing work. This shows that the proposed methodology achieve a lower APL value even with larger ARNET dataset. The successful results in terms of APL have been observed for ARNET larger dataset. To support the APL results of ARNET dataset another big dataset was also incorporated. The results of APL values using this SDFB dataset are summarized in Table 3. The node count for SDFB dataset also ranges from 100 to 900. The APL evaluation for k-APL and APL-proposed are shown in column 2 and column 3.

TABLE 3

APL COMPARISON ON SDFB DATASET

Node Count	k-APL	APL-Proposed
100	4.835277	2.551677
200	3.910016	3.444516
300	4.385205	4.264971
400	7.557384	5.939092
500	7.214229	5.538462
600	6.458818	5.355783
700	4.72845	1.948312
800	4.577479	3.963617
900	6.316097	4.649315

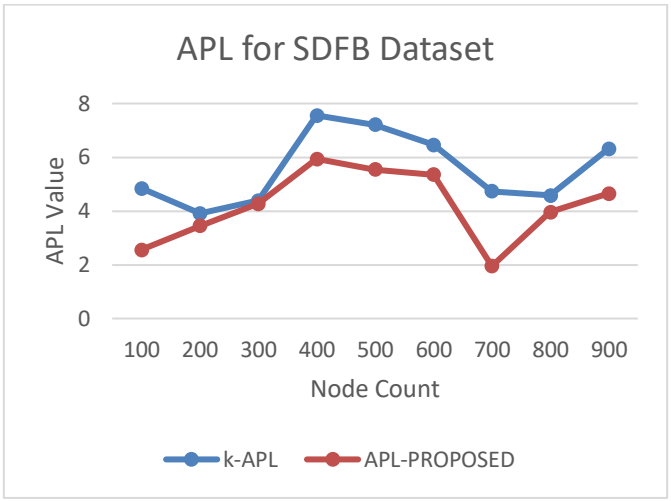

***Fig.7**. APL comparison on SDFB dataset*

Figure 12 shows the APL comparison on SDFB dataset on 900 nodes. It is observed that an overall lower APL values are observed over all the 900 nodes in case of APL-proposed as compared to k-APL. It is also observed that average APL in case of APL-proposed is 4.183 which is lower than the average APL of 5.554 observed in case of k-APL. It can also be understood that APL reduction of 1.371 is achieved by the proposed work over the existing work.

5.2.2 Information Loss Comparison

The proposed architecture design is also evaluated in terms of information loss. Table 4 corresponds to the information loss observed with small ARNET dataset. Table summarized the information loss obtained in case of k-anonymity in column 2 and the proposed work in column 3 corresponding to the 9 nodes as listed in column 1. It is observed that overall information loss of the proposed work is lower than the k-anonymity. The values are plotted in Figure 13 for a graphical comparison.

TABLE 4

INFORMATION LOSS COMPARISON ON SMALL ARNET DATASET

Node Values	Information loss of existing work (k anonymity)	Information loss of proposed work
1	33.0	32.0
2	33.3	33.1
3	33.5	33.0
4	33.6	33.2
5	34.0	33.8
6	34.2	33.1
7	35.0	34.6
8	35.2	34.1

9	36.0	35.8

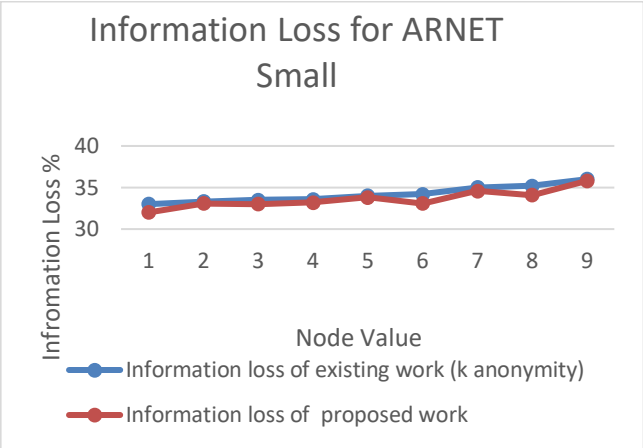

***Fig.8**. Information loss comparison on small ARNET dataset*

The information loss observed against small ARNET dataset is shown in Figure 13. In the graph number of nodes ranging from 1 to 9 is plotted on X-axis against information loss taken on Y-axis. The graph shows that with increase in the number of nodes, there is relative increase in the information loss. It is observed that the average information loss in case of k-anonymity is 34% which is higher than the information loss of 33.63% observed with proposed work. Overall the proposed architecture achieved 0.57% lower information loss due to optimization techniques. Table 5 compares the information loss observed for ARNET big data. The values of information loss observed with existing work on k-anonymity are listed in column 2 and information losses observed with the proposed architecture are listed in column 3. The total number of nodes ranges from 100 to 900.

TABLE 5
INFORMATION LOSS FOR ARNET BIG DATA

Node Value	Information loss of existing work (k anonymity)	Information loss of proposed work
100	33.3848027	28.2381859
200	35.5566881	26.0299738
300	34.0899955	28.0880957
400	35.0781637	21.3738955
500	36.0678233	23.4797667
600	36.8703395	29.927896
700	36.2466593	33.280522
800	36.0056242	25.1810704
900	38.8862447	26.0539086

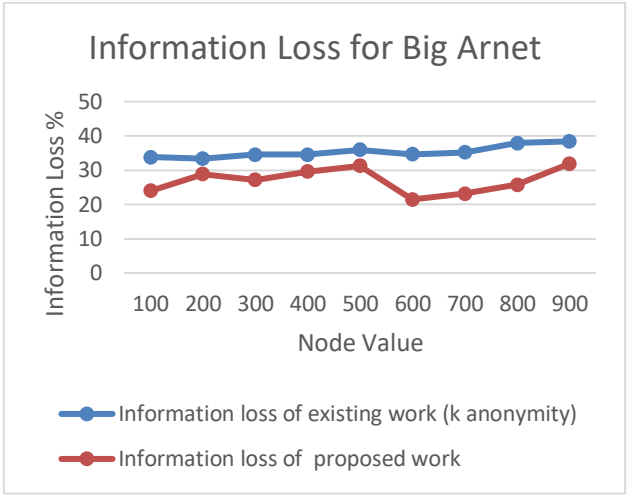

Fig.9. Information Loss for ARNET big dataset

Figure 14 corresponds to information loss observed with ARNET big dataset. The graph shows that information loss of existing k-anonymity shows a slight increase as the number of nodes changes from 100 to 900. Whereas, information loss of the proposed work shows lower values over all the node values. It is observed from the graph that the average information loss of the proposed design is only 26.85% which is much lower than the information loss (35.799%) observed in case of k-anonymity. In other words, the proposed architecture exhibited 8.95% lesser information loss.

TABLE 6
INFORMATION LOSS FOR SDFB BIG DATA

Node Value	Information loss of existing work (k anonymity)	Information loss of proposed work
100	35.9975146	25.980166
200	34.4398286	25.4461521
300	35.1530055	23.364998
400	33.7575702	30.2849738
500	37.0547433	30.9265782
600	37.9913791	26.0495412
700	39.5645466	34.4847617
800	38.8118191	24.2303699
900	37.1416278	28.5873584

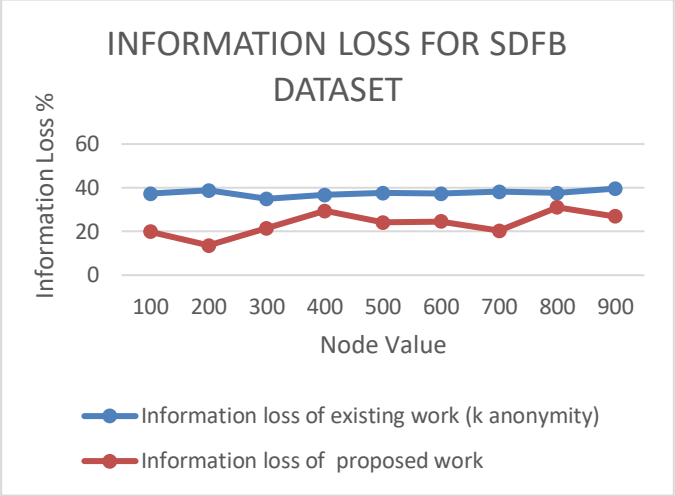

***Fig.10**. Information Loss for SDFB big data*

Information loss was also evaluated against the SDFB dataset. Table 6 summarizes the information loss observed by existing work and the proposed work over the 900 nodes. The information loss values shows that the proposed work exhibits lower information loss as compared to the existing work using k-anonymity. Figure 15 shows the comparison of information loss observed with SDFB big data. It is observed that on average existing work has on average information loss of 36.66% which is very high in comparison to the information loss of 27.71% as observed by the proposed work. In other words, an average the proposed model achieved 8.95% less information loss.

6 CONCLUSION

The proposed design first identifies the sensitive attributes of ARNET and SDFB dataset before optimization performed by the application of ABC and FFBPNN algorithms. The privacy protection was evaluated in terms of average path length and information loss against small ARNET dataset, big ARNET and SDFB datasets. The proposed design involved two fold evaluations by comparing with smaller dataset over 10 nodes and larger dataset over 1000 nodes. APL of the proposed and k-APL using small ARNET dataset are 3.324 and 4.47, using larger ARNET dataset are 3.384 and 5.02 and using larger SDFB dataset are 4.183 and 5.554. Information loss of 33.63% and 34% was observed with small ARNET dataset, 26.85% and 35.799% with larger ARNET dataset and 27.71% and 36.66% with larger SDFB dataset for proposed and k-anonymity, respectively. It is observed that optimization techniques significantly reduced both APL and information loss. This is demonstrated by APL reduction of 1.636 and 1.371 for larger datasets and information loss by 0.57% for small dataset and 8.95% for larger datasets.

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] Keküllüoğlu, D., Kökciyan, N., &Yolum, P. (2016, August). Strategies for privacy negotiation in online social networks. In Proceedings of the 1st International Workshop on AI for Privacy and Security (p. 2). ACM.

[3] 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.

[4] A. Campan and T.M. Truta, "A Clustering Approach for Data and Structural Anonymity in Social Networks," In Privacy, Security, and Trust in KDD Workshop (PinKDD), 2008

[5] Francis, J., & Stokes, M. (2012). U.S. Patent No. 8,140,502. Washington, DC: U.S. Patent and Trademark Office

[6] P. MohanaChelvan 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Tripathy, B. K., Sishodia, M. S., Jain, S., &Mitra, A. (2014). Privacy and Anonymization in Social Networks. Intelligent Systems Reference Library, 243–270

[16] Narayanan and V. Shmatikov, "De-Anonymizing Social Networks", Proc. IEEE 30th Symp. Security and Privacy, pp. 173-187, 2009.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] Zhang, S., Li, X., Tan, Z., Peng, T., & Wang, G. (2019). A caching and spatial K-anonymity driven privacy enhancement scheme in continuous location-based services. Future Generation Computer Systems, 94, 40-50.

[22] 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.

[23] Bhaladhare, P. R., &Jinwala, D. C. (2016). Novel Approaches for Privacy Preserving Data Mining in k -Anonymity Model. J. Inf. Sci. Eng., 32(1), 63 -78

[24] Tsai, Y.-C., Wang, S.-L., Kao, H.-Y., & Hong, T.-P. (2015). Edge types vs privacy in K-anonymization of shortest paths. Applied Soft Computing, 31, 348–359. doi:10.1016/j.asoc.2015.03.005

[25] Chester, S., Kapron, B. M., Srivastava, G., &Venkatesh, S. (2013). Complexity of social network anonymization. Social Network Analysis and Mining, 3(2), 151-166.