RESEARCHERID
THOMSON REUTERS

[Vanitha, 5(12): December2018] ISSN 2348 – 8034
DOI- 10.5281/zenodo.2156480 Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

SEMANTIC DATA ANONYMIZATION USING REINFORCEMENT LEARNING FOR CLOAKING GRAPH PERCOLATION OF SENSITIVE DATA

N.Vanitha[*1] & Dr T. Bhuvaneswari[2]

[*1]Research Scholar, Manonmanam Sundaranar University, Tirunelveli, TamilNadu

[2]Supervisor, Asst. Professor, Department of Computer Applications, Queen Mary's College, Chennai 600004

Abstract

In the age of big data, preserving privacy is a challenging problem to tackle, especially when sharing the graph data generated through social network, users need to share for business analytics and social science research purposes. The top methods among privacy preservation techniques are k-anonymity, I-diversity, differential privacy etc., which prevents re-identification of essential structural nodes in the given graph data. Though the privacy models implemented through such methods may not be completely efficient as the attacker might infer the sensitive data if several nodes of graph database comprises of same labels or attributes. Also, these methods modify the edges between nodes which may significantly alter the essential properties of the database. In this study, we present an algorithm to overcome this challenges with the idea of blocking the graph traversal based on the probabilistic logic to forbid graph percolation which in turn is regulated by reinforcement learning method while ensuring the least amount of distortion in graph properties.

Keywords: *Data anonymization, graph percolation, reinforcement learning.*

I. INTRODUCTION

Social Network data consists of entities represented as either individuals, companies, groups and organizations called as nodes that are joined by one or more properties and connections or links between these entities called edges which indicates some kind of relationship (flows) between these nodes. This flow may carry all sorts of data, so in a social network all the nodes and edges are interconnected. Social Network Analysis (SNA) is a technique used for investigating the mapping and measuring of social structures consisting of either nodes, peoples, groups, organizations and other connected knowledge entities and links representing edges between these nodes. Social Network Analysis is used in geography, information science, sociology etc [1][2]. To perform analysis on Social Network data, data is collected from multiple sources and then data is shared online. Data collected from social networks may contain very confidential and sensitive information about the individuals or users [3]. A social graph usually is acyclic, undirected graph, weighted and these information usually are used by attackers to reveal the identity of the user[4]. The thing that we are facing today is securing the confidential information and takes advantages from Social network analysis. Any information released in Social Network such as Face book, Twitter, etc. that include entities and links between these entities that may lead to privacy implications for involved users. Privacy breach occurs when individuals or organization confidential and sensitive information is disclosed to an adversary [5].Privacy preserving is a method or technique that protects individual and any confidential information in social network [6].So privacy preservation of individuals while sharing individual's collected information in social network is an important research area. Initially the degree of node was considered indicating the number of edges connected to that node with different ones. The enhancement was also there considering the isomorphism, clustering, group formation, changing the structure of graph considered as privacy[7][8].Social network data is usually published with its corresponding relationship intact with one another. In principle once anonymized, the non-sensitive labels or attributes are used to recall the identifiers to recover the sensitive information form tabular micro-data [9][10]. For example: the anonymized social network data acquired from two sources i.e., Zomato (a restaurant discovery service website) and Book My Show (an online movie booking portal) could be used by

attacker to find co-relation of the user's identity and place based on the relation between several labels which points towards user's facebook account thereby comprising user's privacy. As these two websites heavily depends on facebook database.

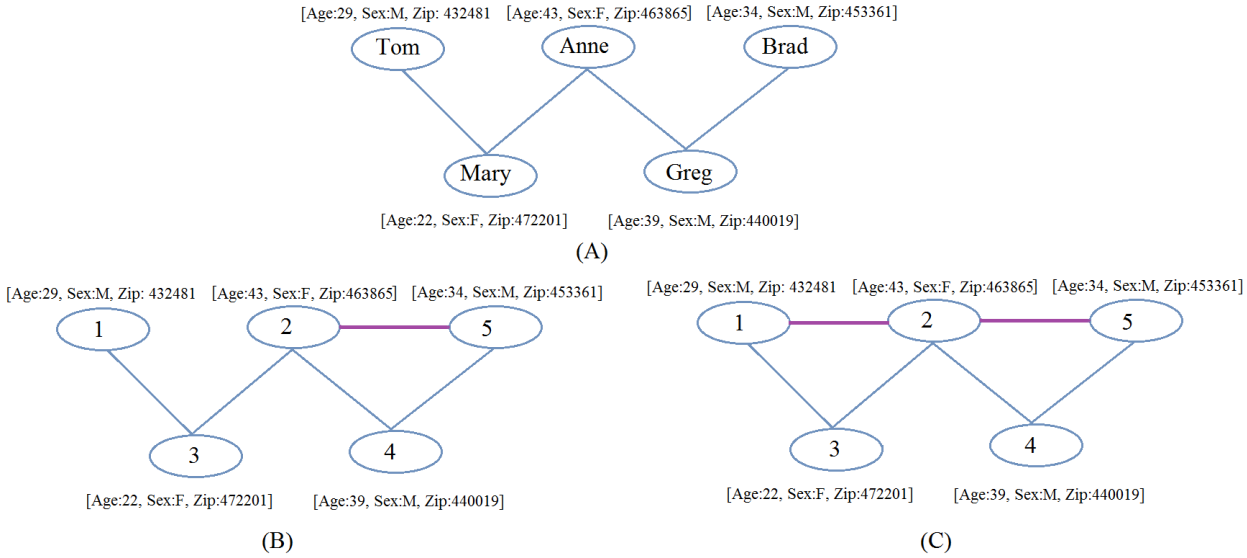

Figure 1: (A) Illustration of original social network based graph data, (B) 2-Degree Anonymous Graph, (C) Anonymous graph which satisfies 2-degree and 2-diversity graph.

A structure attack is directed over degree and sub graph in order to identify the node. Thus, in order to prevent a structure attack the anonymized graph should satisfy k-anonymity [11-14]. Thus, by increasing this degree of relation and diversifying the connection between edges and vertices would ensure in preserving anonymity as shown in figure 1(A), (B) and (C). In this context the research work done by Liu and Terzi had pioneering accomplishments to define k-degree based model of anonymity to prevent degree attacks [11]. Here, for any node in the graph if it has at least k nodes with same degree is said to be k-anonymous. For example if an attack is aware of the node 4 being related with 2 other friendly nodes then it can be immediately inferred the identity of the related nodes based on sets of related attributes. Thus, k-anonymity is an essential tool to fend off such attacks [15, 16]. But k-anonymity itself is not a comprehensive solution to curb the risk of compromising privacy. Theoverall approaches available for protecting privacy can be classified into two sets i.e., clustering and edge splitting.

1. Clustering: This method involves merging sub graphs to singular sub-node thereby leading to loss of all node-label relations though it is unsuitable for sensitive labels in a graph database.
2. Edge Editing: Another approach is edge editing, as the name suggest the node-label relation is altered by either swapping/adding/deleting the edges though the original node relation remains unchanged.

To address this issue, we propose a probabilistic model of reinforcement learning to dictate the levels of graph percolation. Graph percolation is heavily used in statistics, physics and mathematics to portray the behaviours of connected clusters or complex networks in a graph. As the data extracted from social network follows power law distribution. Therefore the existence of low degree graph and its connections could be logically used to dictate the path of the relations between several nodes or to hide noises from repeatedly getting re-spotted with sensitive attribute.

ISSN 2348 – 8034
Impact Factor- 5.070

II. LITERATURE REVIEW

Survey of this paper proposes quantitatively to achieve anonymization for showing the semantic cloaking and labelling, which can develop the privacy based mobility dataset can be replaced by the semantic categories. In order to improve the framework of semantic labelling to evaluate the dataset uniqueness (ϵ)measures[17].Most of the data can be represented as graphs, with real world entities as graph nodes and interrelationships among entities as graph edges. Mining these released data, or corresponding graphs, may facilitate the forming of judicious strategies for marketing or promoting public health. However, individual data inevitably contain private information. How to prevent potential adversaries from recognizing the mapping between a particular graph node and a real world individual is critical for data providers. This paper proposes a semantic-based data anonymization method which employs entity ontology to anonymize the graph data for publication [18].Social network data are publicly available, analysed and utilized in one or another way since it leads to an important issues in privacy preservation. This paperproposes the existing technique of anonymization approach based on social network data for privacy preserving. This problem formulationis done by using data utility, knowledge and privacy as the dimensions [19].This paper provides greater privacy than greedy perturbation technique in social network analysis. Privacy preservation has a trade-off between the utility of data and preservation of sensitive information. This process improves with minimal concerns the privacy of sensitive information to utility [20].This paper is motivated by the recognition of the need for personalized privacy and finer grain in data publication of social networks. Recently, investigators have proposed a privacy protection system that not only avoids the discovery of identity of users but also the discovery of particular features in user's profiles[21].The need of improving the privacy on data publisher becomes more important because data grows very fast. Privacy preserving data publishing is traditional methods which cannot be preventing privacy leakage. This will causes the research to find better approaches to prevent the privacy leakage. The well-known techniques of K-anonymity and L-diversity are mainly used for data privacy preserving. These techniques on the data privacy cannot prevent the similarity attack since they did not take into semantic relation between the sensitive attributes of the categorical data. In this paper, we proposed an approach to categorical data preservation based on Domain-based of semantic rules to overcome the similarity attacks [22].

III. METHODOLOGY

2.1.Experimental Dataset

The proposed model is implemented using MATLAB R2012a under Windows platform. The experiments are conducted over the machine with hardware configurations of Intel's third generation 8-core microprocessor with NVidia 630 graphic card, 2GB RAM giving a fine clocking speed of 2.7 GHz. The consolidated databases used in the study are from Arnet Data Set, Cora data set, and DBLP data set. Mostly citation based public database has been used for the experimentation and performance analysis. The detailed properties of the used graph data are represented in the table 1 below.

Table 1 Types of Database Used

S.I No	Database	Nodes	Edges	Domain
1	Arnet Data Set [23]	6,000	37,848	Citation Network
		6,000	37,848	Advisor-Advise Network
2	Cora data set [24]	2,708	5,429	Citation Network
3	DBLP data set [25]	6,000	29,843	Citation Network

2.2 Model

To start with the graph percolation model, first we create a graph model. Let us suppose that a given radial social network data comprises of b+1 number of vertices. Thus, it can be modelled with the help of a graph tree, which is given as G =(N, E) where N represents the sets of vertices i.e, N=1,...,N_n and E is the edge set with the cardinality is given as $|N| = E$. Here, each of the edges in form of the tree is rooted for index value n ruled by the probability flow from one edge to another and is derivable from source S to destination D with j amount of deviation in the form

S_n+jD_n. Then the whole branch probability based graph flow model can be derived by initialising the conditional probability in form of a sequence traversed from parent to child nodes of two local posteriors i.e, probability of appearance of vertices (P_1) and that of edges (P_2) which is given by: $P_1 = (N^t|E_1^{1:t})\&P_2 = (N^t|E_2^{1:t})$. This is represented in the form of sequencized finite sets with multi object densities of $E_i^{1:t}$ observed edge sites. Here, the synchronization between such posterior is maintained as:

$$P_\alpha(N^D|E_1^{S:D}, E_2^{S:D}) = P_\alpha(N^D|E_1^{S:D} \cup E_2^{S:D}) \quad \text{.....(1)}$$

Now, to overcome the problem of unknown correlation between no two distributions of independent variables the solution is:

$$P_\alpha(N^D|E_1^{S:D}, E_2^{S:D}) \propto \frac{P_\alpha(N^D|E_1^{S:D})P_\alpha(N^D|E_2^{S:D})}{P_\alpha(N^D|E_1^{S:D} \cup E_2^{S:D})} \quad \text{.....(2)}$$

Hence, the generalized posterior relationship can be represented in the form of geometric mean:

$$P_\alpha(N^D|E_1^{S:D}, E_2^{S:D}) = \frac{P_\alpha(N^D|E_1^{S:D})^{\alpha 1} P_\alpha(N^D|E_2^{S:D})^{\alpha 2}}{\int P_\alpha(N^D|E_1^{S:D})^{\alpha 1} P_\alpha(N^D|E_2^{S:D})^{\alpha 2} \delta N} \quad \text{.....(3)}$$

Where, $\alpha 1, \alpha 2$ $(\alpha 1 + \alpha 2 = 1)$ the parameters determining the relative probability of weighted distribution (w) between a specific hierarchical level of child and parent nodes. Now, in order to anonymize the percolated model of graph data, we need a rule set to algorithmically eliminate the sensitive vertices or add extra edges between labelled vertices to disrupt the probability of finding sensitive information. In this way we can induce anonymity in a quantitative way over the graph database. To do so, we have used reinforcement learning (RL). RL functions based on the penalty and reward system for executing operation i.e, if the operation sequence is executed suitability then the RL algorithm is rewarded else penalty is induced in it. Thereby, leading the results towards more convergent solution. Another advantage of using this method is that the algorithm can form the ideal strategy of adding or removing vertices and edges respectively in spite of unavailability of prior information or trainable data.A strategy for an addition of edges and removal of sensitive vertices is assigned using RL upon the above modelled graph percolation data, at each time *t*, for each statesa probability for performing action $a \in U(s)$, as per the given history as:

$$H_{t-1} = \{s_1, a_1, r_1, \dots, s_{t-1}, a_{t-1}, r_{t-1}\} \quad \text{.....(4)}$$

This incorporates the states, actions and rewards observed until time $t-1$. A policy P is taken in account to determine the reward or penalty for a particular action an against state s. P is memory-based technique, i.e., it primarily depends only upon the memory of the history of state and not onto its current state. Thus, a deterministic strategy P assigns each state a unique action a. While taking after a strategy P we perform at time t action a_t at state s_t and observe a reward r_t (distributed according to $P_{GP} = P_\alpha(N^D|E_1^{S:D}, E_2^{S:D})$)and the next state s_{t+1} (dispersed according to $P_{S_t,S_{t+1}}^{P_{GP}}(a_t)$). Here, action a corresponds to addition of edges or removal of vertices. We consolidate the sequences of rewards to a single value called the return, and our goal is to minimize the probability of discovering sensitive information by manipulation the graph connectivity of previously marked sensitive labels to vertices. Hence, we concentrate our work to focus on discounted return, which has a parameter $\gamma \in (0,1)$, and the discounted return of policy P is:

$$P_{P_{GP}}^P = \sum_{t=0}^{\infty} \gamma^t r_t, \quad \text{..... (5)}$$

Where r_t is the reward observed at time *t*. Since all the rewards are bounded by P_{Max} the discounted return is limited by:

$$P_{Max} = \frac{P_{Max}}{1-\gamma}. \quad \text{.....(6)}$$

For a sequence of pairs for state and action, let the covering time, denoted by C', be an upper limit on the number of state-action pairs beginning from any pair, until all state-action appears in the sequential arrangement. Note that the covering time can be a function of both the P_{GP} and the sequential arrangement or just of the sequence itself. This previously mentioned policy $P_{P_{GP}}^P$ generates the sequence of state action pairs. The upside of this model is that it

permits to ignore the exploration and to concentrate on the learning. In some sense P($P^{P}_{P_{GP}}$) can be viewed as a flawless exploration approach. The following equation of RL algorithm gauges the state-action value function as takes after:

$$Q_{t+1}(s,a) = \alpha_t(s,a)\left(R_{MDP}(s,a) + \gamma \max_{b\in U(s')} Q_t(s',b)\right) + \left(1-\alpha_t(s,a)\right)Q_t(s,a) \quad \text{.....(7)}$$

Algorithm: ***Reinforcement Learning based Cloaking Graph Percolation in Semantic Data***

Input: Graph tree, which is given as: G:=(N,E); where N & E represents the sets of vertices and sets of edges. P_α *is the probability tree of graph percolation for the above graph database.* s_i,s_j *are the probability division of* P_α *derived from RL based on parent and child node hierarchy,* S_t *is the Graph Splicing Vector at instance t,* H_t *is the history of state –action pairs at time t &* p_t *is the pair equilibrium sequence.*
Output: H_o *is the online hypothesis & spliced graph comprises of new* P_α *i.e,* P'_α
Step 1: For I=1, 2, 3 …, N

Step 2: Receive New Instance
Load: $H_t = \{s_1, a_1, r_1, \dots, s_t, a_t, r_t\}$
s_i,$s_j \in Q(s,a): P_\alpha$ .....(8)

Step 3: Form pairing between two hierarchical nodes
$p_t = \frac{[s_i * s_j * q]}{[s_i * s_j]}$ $//paired\ equilibrium\ sequence$ .....(9)
$S_t = \sum_i p_t\,(t-t_i)$ $//\ Graph\ Splicing\ Vector$ .....(10)

Step 4: Evaluate risk of recovery of sensitive label:
While $k > S_i \&\& k \neq D_i$
{
$R[k_i] = \frac{1}{p_t}\sum_{l=1}^{t}(S_t - S_{t-1})$ $//risk\ of\ recovery$ .....(11)
$K_t = \sum_k k\,(t-t_i)$ .....(12)
}
$where, k = \frac{(p_t\,\Delta t)^k}{k!}\exp(-p_t\,.\,\Delta p_t)$ .....(13)

Step 5: Evaluate thresholds of the graph splicing matrix:
If $P_{Max} > 0$
{

$$\sum_{i=1}^{k}\rho_{N,E} = \begin{cases} 1 & if\ S_t \geq \frac{\sum_{i=0}^{k} w}{\sum_{j=1}^{c'} l}\ //remove\ vertices \\ 0 & if\ S_t < \frac{\sum_{i=0}^{k} w}{\sum_{j=1}^{k} l}\ add\ edges\ connecting\ risk\ free\ labels \end{cases} \quad \text{.....(14)}$$

w is relative probability of weighted distribution between a specific hierarchical level of child and parent nodes. Also, the k is the number of iteration in computation.

else
break;
}

Step 6: Update the hypothesis with chain sequence by checking for overall risks involved in the current hierarchical level:
if $R[k_i] \leq R[S_t]$

ISSN 2348 – 8034
Impact Factor- 5.070

{

$$P'_{\alpha} = \sum_{k=1}^{Si} \sum_{k=1}^{Di} \frac{(p_t)^{k-k_i}}{(R[k_i]-t_i)!} . \exp(-p_t \ . \ \Delta t) . \frac{(p_t)^{k_i}}{(k_i)!} \quad \text{.....(15)}$$

Update the state and actions using:

$$Q_{t+1}(s,a) = \alpha_t(s,a)\left(R_{MDP}(s,a) + \gamma \max_{b \in U(s')} Q_t(s',b)\right) + \left(1-\alpha_t(s,a)\right)Q_t(s,a) \text{.....(16)}$$

Return P_{Max}

Update:

$$H_{t+1} = \{s_1, a_1, r_1, \dots, s_{t+1}, a_{t+1}, r_{t+1}\} \quad \text{.....(17)}$$

}

else

{

Print "Failed to Update"

$$k_i = k_{i+1}$$

Return:

$$H_{t-1} = \{s_1, a_1, r_1, \dots, s_{t-1}, a_{t-1}, r_{t-1}\} \quad \text{.....(18)}$$

}

Step 7: End process

The main advantage posed by the above algorithm is the setup it put forth for scheduling training and processing based on memory bound to the support set on timely basis and thus forming a decomposable or expanding sequence when the other instance pairs are added in relational to the previous trained hypothesis, such that the trained hypothesis is always bounded and deducible from the other previous pairs of instances. The training is achievable in small number roof instances with high accuracy. Here, the figure 2 (A) represents the onset of the data slip rate due to the bounded memory units to update the online hypothesis. However, the end of the recovered image the data slip rate haven't shown any fluctuation in loss of data, thereby converging the results at the saturation level with respect to the time see fig 2(B). The advantage that our online learning algorithm put forth is its adaptability in re-organizing the online hypothesis so generated during the process run. This lead to a curvelet transformation for the process to end at the point where the front data and the half pattern length saturates. The process is cyclic in nature and doesn't require unnecessary updating process.

ISSN 2348 – 8034
Impact Factor- 5.070

IV. RESULTS & DISCUSSION

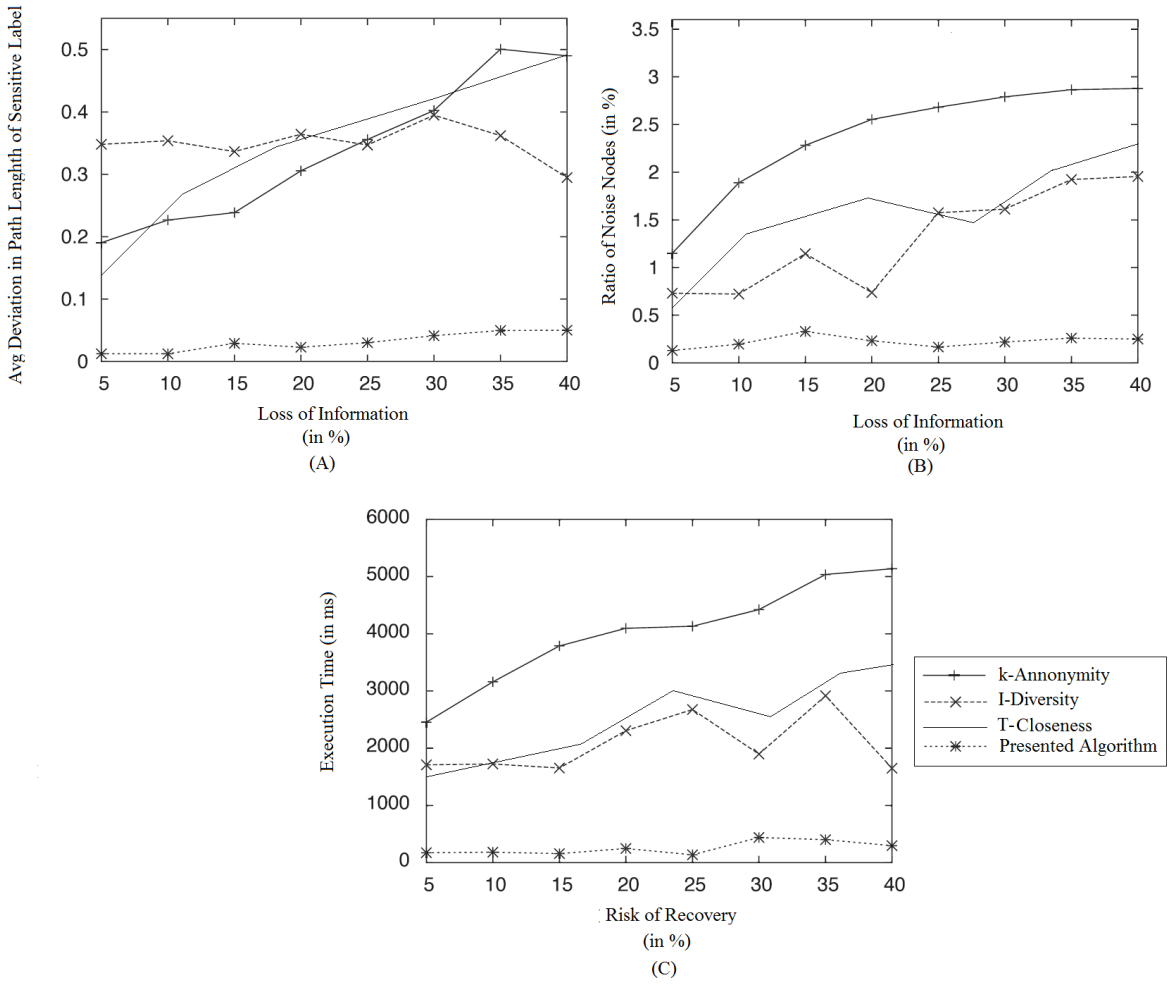

Figure 2 Performance Comparison of the presented method with other methods for datasets (A) Arnet Data Set, (B) Cora Data Set and (C) DBLP Data Set

In order to exhibit the effectiveness of the presented method, we have compared the results from our work with the other most prominent methods such as: k-anonymity, I-Diversity, T-Closeness. The above figure 2 (A) represents the performance results for Arnet data set. Here, the average deviation in path length of sensitive labels in an anonymized graph is compared with the loss of information. As we can infer from the graph that as the length of the path increases the loss of information increases owing to the increased sparsity between relationships between anonymized vertices. Although the performance dip can be noticed in other methods ranging from 0.1-0.2 ratio but the presented method shows convergent results with deviations ranging from 0.03-0.05, this is due to the nature of algorithm which preserves the distance between nodes by quantitatively analysing the optimal reorganization or manipulation of graph required to preserve privacy. Additionally, from the figure 2 (B) the loss of information is also resulted to be minimal when comparing with other methods. Since, the conventional methods clocks higher computational time owing to its data intensive computation but in the presented method since we are computing the reorganisation of graph based on probability flow of percolated graph thus drastically reduces the computational expenses require executing the program. From the figure 3(C)represents the performance results for DBLP dataset. Here, the execution time in an anonymized graph compared with the Risk of Recovery. As we can infer from the graph that Risk of Recover is resulted to be minimal when comparing with other methods. Execution time is

minimized in presented algorithm when compared with K-annoymity, I-diversity and T-Closeness. All though the performance of Execution time can be noticed in other methods ranging from 1500-2500ms but the presented method shows the execution time is 0.03ms,this is due to nature of algorithm which preserves the execution time thus drastically reduces.

In this study we haven't discussed the concept of swapping the sensitive labels as it is a computationally expensive process and more often there are few cases where applicability of this method would find suitable grounds of validity. Here the RL algorithm functions on state and action relationship pairs thus because of the modelled policy the noise states gives zero return and thus filtered out from the sequence with higher rewards, thereby eliminate the influence of noise nodes contribute in output. For higher protection the k-neighbourhood method follows the principle of isomorphism i.e, for every graph there lies a minimum of k+1 nodes in such cases it is ensured that if the attacker find the sensitive labels it will have at least two possible candidates fulfilling the scenario but in such cases the method increases the size of the graph database and thus render it useless for data analytics. As more edges and vertices means more computational time in graph traversal making the problem to push towards NP-Completeness with no feasible algorithm to perform analytical study over it. A possible solution is to anonymize the graph by blocking the graph traversal path leading to maximum likelihood of discovering information for a tree search. This is readily achieved in our method and that too without increasing the size of the graph, thus making it more feasible option for large scale social network data analysis.

V. CONCLUSION

In this study we presented a reinforcement algorithm based probabilistic method for blocking the graph traversal by regulating graph percolation. The rigours analysis of the presented algorithm gives satisfactory results within the theoretical bound of noise nodes added while ensuring the properties of the graph database remains intact. The experimental results shows that our modified version of edge editing has superior advantages than compared with past methods while eliminating the need to crowd the graph with noise nodes by quantitatively learning the sequence of operations required to achieve optimal edge editing scenarios. In these scenarios where publishers publish their data in a distributed environment the attacker can still co-relate the data and its subsequent labels but with the graph percolation protocols the implementation of the presented method gives satisfactory results in hindering the co-relation of distributed datasets.

REFERENCES

1. *Vedanayaki M. A study of data mining and Social Network Analysis. Indian Journal of Science and Technology. 2014 Nov; 7(S7):1–3.*
2. *Mittal P, Garg S, Yadav S. Social Network Analysis using interest mining: A critical review. Indian Journal of Science and Technology. 2016 Apr; 9(16):1–8.*
3. *Zhou B, Pei J, Luk W. A brief survey on anonymization tech- niques for privacy preserving publishing of Social Network Data. Association for Computing Machinery SIGKDD Explorations Newsletter. 2008; 10(2):12–22.*
4. *Rajper S, Shaikh NA, Shaikh ZA, Mallah GA. Automatic detection of learning styles on learning management sys- tems using data mining technique. Indian Journal of Science and Technology. 2016 Apr; 9(15):1–5.*
5. *Singh A, Bansal D, Sofat S. Privacy preserving techniques in Social Networks Data Publishing - a Review. IJCA. 2014; 87(15):1–6.*
6. *Hariharan R, Mahesh C, Prasenna P, Kumar RV. Enhancing privacy preservation in data mining using cluster based greedy method in hierarchical approach. Indian Journal of Science and Technology. 2016 Jan; 9(3):1–8.*
7. *Masoumzaden A, Joshi J. Preserving structural properties in edge-perturbing anonymization techniques for Social Networks. IEEE Transactions on Dependable and Secure Computing. 2012; 9(6):877–89.*
8. *Nandi G, Das A. A survey on using data mining tech- niques for Social Network analysis. International Journal of Computer Science Issues. 2013; 10(6):1–25.*
9. *G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "Fast Data Anonymization with Low Information Loss," Proc. 33rd Int'l Conf. Very Large Data Bases (VLDB '07), pp. 758-769, 2007.*

10. *G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "A Framework for Efficient Data Anonymization Under Privacy and Accuracy Constraints," ACM Trans. Database Systems, vol. 34, pp. 9:1-9:47, July 2009 .*
11. *K. Liu and E. Terzi, "Towards Identity Anonymization on Graphs," SIGMOD '08: Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 93-106, 2008 .*
12. *B. Zhou and J. Pei, "Preserving Privacy in Social Networks Against Neighborhood Attacks," Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE '08), pp. 506-515, 2008 .*
13. *J. Han, Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, Inc., 2005 .*
14. *L. Zou, L. Chen, and M.T. O ¨zsu, "K-Automorphism: A General Framework for Privacy Preserving Network Publication," Proc. VLDB Endowment, vol. 2, pp. 946-957, 2009.*
15. *A. Campan, T.M. Truta, and N. Cooper, "P-Sensitive K-Anonymity with Generalization Constraints," Trans. Data Privacy, vol. 2, pp. 65-89, 2010 .*
16. *B. 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, pp. 47-77, 2011.*
17. *OmerBarak,Gabriellacohen and EranToch, "Anonymizing mobility data using semantic cloaking" ,2015 Elsevier.*
18. *Shu-mingHsieh,Mao-Hsuyen and Li-jenka, "Semantic-based graph data anonymization for big data analysis", 2016,International Conference on Machine Learning.*
19. *V.VijeyaKaveri and Dr.V.Maheswari, "Cluster Based Anonymization For Privacy Preservation in Social Network Data Community" ,March 2015,Journal of Theoretical and Applied Information Technology,Vol.73,No.2.*
20. *NayanMattani, J. Sharath Kumar, A. Prabakaran and N. Maheswari , "Privacy Preservation in Social Network Analysis using Edge Weight Perturbation" , October 2016,Indian Journal of Science and Technology, Vol 9(37), DOI: 10.17485/ijst/2016/v9i37/93810.*
21. *Mr.Gaurav .P.R. and Mr.Gururaj.T, "Anonymization: Enhancing Privacy and Security of Sensitive Data of Online Social Networks" ,2014, International Journal of Computer Science and Information Technologies, Vol. 5 (4).*
22. *Ahmed Alimubark, Emad Elabd and Hatem Abdulkader , "Semantic anonymization in publishing categorical sensitive attributes" ,24 March 2016, IEEE.*
23. *Arnet Data Set, https://aminer.org/data*
24. *Cora data set, https://relational.fit.cvut.cz/dataset/CORA*
25. *DBLP data set, http://dblp.uni-trier.de/*