ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Caring Perceptive Labels in Social Network Data Anonymization

J.Manjusha
M.Tech Student,
Department of computer science & Engineering,
Madanapalle institute of Technology & Science.

S.Murali Krishna
Professor & Head of Department,
Department of Computer Science & Engineering,
Madanapalle institute of Technology & Science.

ABSTRACT:

In recent years, the rapid growth of web applications developed the need for private data to be published. Most of the social network data necessitates the data to be available for easy access and conversion of data to graph structure to re-identify sensitive labels of individuals became an impeccable issue.In this paper, we have made a detailed surveyed about the existing techniques that preserve the sensitive data in social network.It is observed that preserving the graph structure and label re-identification by adding some noise nodes to the graph makes significant change in degree is inferred from existing techniques. The anonymization methods for preservation of the private sensitive data based on cluster based approach and graph modification approaches are studied in detail.

Index terms:

Graph Modification Approach, Data Anonymization approach, k-Anonymity Model,Verification loss.

I.Introduction:

The social network is useful in social sciences and researches to study the relationship between individuals or among societies. There are various types of social networking websites that connects the individuals to share the data. The relationships between the individuals are mentioned by a graph. The graph constructed for these online communities and social networks are very complex. While sharing the data or information the graph will also be published along with the node, identities can be removed. A survey in June 2012, US demographics says the age distribution in social networks and online communities in average of 24 sites. They surveyed that teens and youth peoples are the heaviest users of social networking sites.

Specifically, 16% of peoples around 18-24 years, 26% of peoples around 25-34 years and 25% of peoples around 35- 44 years using the social network and online communities. So more number of data has been shared and the relationship graphs for individual users are more complex. Even though they are preserved, the intruders could use to apply the security breaching algorithms for the vulnerable nodes and grab the information of individuals.

The social networking sites are the service providers used to connect with peoples to share the information like photos, videos and personal messages. As the social networks usage grows, the risks behind them also increases. That makes the hackers, spammers, virus writers to attack the vulnerable nodes. An intruder may infringe the privacy of the node with the help of published social network data and some background knowledge. For better privacy, the identities of the label such as social security number (SSN) of an employee, disease of a patient, etc., are replaced by some unique identity.

II.Literature survey:

In a social network, peoples are represented as nodes and the link between the nodes are represented as edges. Using passive and active attacks, adversaries may notice the edges between specified targeted pair of nodes in the copy of original graph. Lars Backstrom, Cynthia Dwork and Jon Kleinberg (2007) [1] described about the attacks that the adversaries can identify the target nodes by adding some new nodes to the sub graph of the anonymized graph. The attacks that the attackers used to identify are, In attackers' point of view, adding some new nodes randomly to the graph, it will not receive any messages from other nodes. When publishing the graph, attackers need to find the copy of the fake nodes and named

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

it as x1,x2,..xn and also find the original nodes and named it as w1,w2,….wn. So the privacy has been compromised. If the graph is directed, then the attackers become easier so they focused on undirected graph. In the undirected graph case, there is an edge (u, v) if at least one of the directed edge (u, v) or (v, u) is present. The nodes are added is O (log n) it is higher band. Here, at least Ω (log n) nodes are needed for active attacks. In walk-based attack, have efficient algorithms. It is harder to detect the data for data holder.

In cut based attacks, it produces densely connected component attached weakly to rest of the graph. Concluding that, the results are forced by impossible results when computational defends to protect the social network data. So, moving to the individual privacy and permit accurate analysis. SmritiBhagat, Graham Cormode, Balachander Krishnamurthy and DiveshSrivastava(2009)[2] discussed about the interactions between the targeted node. They grouped the individuals into classes and cover the relationship between nodes in anonymized graph. To prevent the inferences in interactions of nodes, critical safety conditions are formed in those classes which are grouped.

SmritiBhagat, Graham CormodeBalachander Krishnamurthy and DiveshSrivastava(2009)[2] present the —label list approach, in this approach, the labels are given to each node that identifies the list of possible identifiers includes its true identifiers. Those lists are structured and the true identifier cannot be inferred and the pattern of links between classes doesn't leak information. An attacker correctly guess the entities who participated in interactions by using the published graph which is in the form of (k,m) uniform list approach. If the graph published as full list, an attacker can identify using the background knowledge. If an attacker has complete knowledge about one node and partially knows about its relational nodes, using the interactions between those nodes, the attacker can identify the information of particular node. To prevent this attack, in which influence the information of individuals increase the amount of masking the data. So moving to partitioning approach. Here only the edges between subsets are released rather than releasing full edge information. The dataset employed is Xanga social network dataset consists of 780 nodes and 3 million edges and Speed Dating dataset which consist of 530 participants and 4150 dates.

For Xanga dataset, full list approach set to k=m=10 and prefix list approach set to k=10,m=20 guarantees individual privacy with probability at least 90% than the previous studies. It requires less than 1 minute to anonymize, operates with 1GB of RAM. For Speed Dating dataset, parameters are similar k=m=5 for full list and k=5, m=10 for prefix list approach. The relative error exceeds 100% than query results.

The Privacy breaches occurs more in the social network data. The anonymization techniques made the privacy for individuals in the published anonymized graph. AlinaCampan and Traian Marius Truta (2008)[3] developed the Greedy privacy algorithm for anonymize the social network and introduces the structural information loss measure. The privacy model uses only binary relationship. K-anonymity is the best model for microdata privacy protection. Each and every identifier is indistinguishable with other individual identifiers. The edge generation is processed with two components.

1.Edge inter cluster generalization:

Social Network GREEdy Algorithm (SANGREEA) [3] generates k-anonymous masked social networks. The quasiidentifiers sensitive attributes are partitioned into clustersand made the relationship between them. Input to the algorithm is a graph, G=(N,E) and k- as in k-anonymity.α,β – user defined weight parameters. Theoutput of the algorithm is as : S= {cl1,cl2,….,clv} – A set ofclusters.Authors conducted the experiment by running theSANGREEA algorithm for set of data, k=3,α=1,β=0 andk=3,α=0,β=1. Comparing this algorithm with Zheleva'salgorithm. Both are implemented in java, executed on dualCPU machine with 3GHz and 4gb RAM.

Experiment isperformed for social network with 300 nodes selected inADULT dataset from UC Irvine machine repository Theinformation loss is smaller than Zheleva's algorithm.All the data privacy models focus on masking the dataand protecting quality of it. The novel approach is proposedby AlinaCampan, Traian Marius Truta and Nicholascooper (2010) [4] for safeguarding the usefulness of initialdata that deformation performed. This paper discusses theprevention of the attribute disclosure against thegeneralization constraints and privacy models. This definesthe preferences for numeral attributes in terms of ranges.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

AlinaCampan, Traian Marius Truta and Nicholas cooper(2010)[4] focuses on privacy and utility of constraints intransactions and on k-anonymity.The algorithm used is GREEDY CPKA. Where the inputto Algorithm is Initial micro data (IM), p,k as in psensitivek-anonymity generalization boundaries. Thealgorithm works and generates a set of clusters S= {cl1,cl2,... .,clv,clv+1}. The quality of results measured by the-normalized total information loss metrics. The algorithmefficiencies are articulated by their running time. Data setused id adult dataset from UC Irvine machine learningrepository Consists of 45,222 tuples. Information loss\ doesn't degrade drastically when constraints areincorporated into p-sensitive k-anonymity.

Graham Cormode, DiveshSrivastava, Tinhyu and Qing-Zhang (2008)[5], introduced (k,l)-groupings. It safeguardsthe graph structure perfectly. It guarantees to resist theattacks. It is used to find the safe groupings. This (k, l)-groupings offer privacy-utility swapping. The experimentsare done on DBLP dataset, to evaluate the usage ofanonymized data. Graham Cormode, DiveshSrivastava,Tinhyu and Qing Zhang (2008)[5], studied the graphs asbipartite graph data. The graph is thin and must make surethat the associations are not exposed. The graph structureshould be protected rather than masking the graphstructure. Grouping of nodes forms a NP-Hardness problemfor safe grouping. The problem in this paper is, the bipartitegraph is not exactly anonymizing the data.

III.Existing work:

The present trend social networks does not giving the security about the user profile views. In the existing systems the data distribution takes more time and not perfectly displaying sensitive information and non sensitive information.

Drawbacks

» There is no choice to publish the non sensitive data to all.

» There is no security to user profiles.

» Easy to leakage the private information and attacks by the malicious adversaries.

IV.Proposed system:

We overcome above problems proposingThe anonymization methods for preservation of the private sensitive data based on cluster based approach and graph modification approaches. These are observed that preserving the graph structure and label re-identification by adding some noise nodes to the graph makes significant change in degree is inferred from existing techniques.This anonymization preserves the privacy of individuals but it would cause the graph to be useless for any study. So, to overcome this, propose some additional requirement that minimizing the edge modifications is done. So the usage of original graph is preserved, at the same time the degree anonymity constraint also satisfied.

V.SENSITIVE DATA ANONYMIZATION APPROACHES:

To preserve the privacy attacks in data publishing, data should be anonymized properly before the publishing of data. Generalization and perturbation are the two anonymization approaches for relational data. Some other categories of anonymization methods are discussed in the forthcoming inferences.

A.Cluster Based Approach :

The cluster based method groups the similar vertices into a group and forms a sub-group. Similarly clusters the edges and forms a sub-group [17]. And finally anonymizes the sub-groups into a super vertex. Thus the information about the individuals can be hidden properly. The cluster based approach is categorized into, vertex clustering methods, edge clustering methods, vertex and edge clustering methods and vertex-attribute mapping clustering methods.

B.Graph Modification Approach :

The graph modification approach [17] anonymizes the graph by modifying the edges and vertices of the graph. The modifications are conducted by three ways namely.

1. The optimization approach
2. The randomized graph modification approach
3. The greedy graph modification approach

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

C.Edge Editing Approach :

The edge editing approach [17] will not change the nodes in the original graph but modifies (add/delete/ swap) the edges. It destroys the properties of graph by adding a new node to it. It will change the distance properties by connecting the two faraway nodes by linking between those two communities. So preserving the usage of data is not providing the good solution. By carefully adding the noise nodes to the graph may provide preservation of some graph properties. The edge editing method follows the neighbourhood rule to modify the edges.

D. k-Anonymity Model :

The k-anonymity [13] model protects the structure attacks and protects the re-identification of the nodes. A data holder shares ones private information to researchers that can be re-identified by the attackers. So the k-degree anonymity model is proposed.
The attacks against k-anonymity are,
1. Unsorted matching attack
2. Complementary release attack

The disadvantage in this paper is that, this protection model has some attacks. So this protecting privacy model is not satisfied for individual privacy.

E. l-Diversity :

The l diversity model [12] prevents the attacks of k-anonymity. This model is more secure than k-degree. By adding the edges between the nodes which are have l-distinct values. The nodes have some degree but the labels are different. So, the attacker cannot conclude the information of a specific node. It prevents the homogeneity attack and background knowledge attacks and it is implemented efficiently.

VI.TECHNIQUES FOR PRIVACY PRESERVATION OF SENSITIVE DATA IN SOCIAL NETWORKS :

In a social network, peoples are represented as nodes and the link between the nodes are represented as edges. Using passive and active attacks, adversaries may notice the edges between specified targeted pair of nodes in the copy of original graph.

Described about the attacks that the adversaries can identify the target nodes by adding some new nodes to the sub graph of the anonymized graph. The attacks that the attackers used to identify are,

A. Walk-Based Attack:

In attackers' point of view, adding some new nodes randomly to the graph, it will not receive any messages from other nodes. When publishing the graph, attackers need to find the copy of the fake nodes and named it as x1,x2,..xn and also find the original nodes and named it as w1,w2,….wn. So the privacy has been compromised. If the graph is directed, then the attackers become easier so they focused on undirected graph. In the undirected graph case, there is an edge (u, v) if at least one of the directed edge (u, v) or (v, u) is present. The nodes are added is O (log n) it is higher band.

B. Cut –Based Attack:

Here, at least Ω (log n) nodes are needed for active attacks. In walk-based attack, have efficient algorithms. It is harder to detect the data for data holder. In cut based attacks, it produces densely connected component attached weakly to rest of the graph. Concluding that, the results are forced by impossible results when computational defends to protect the social network data. So, moving to the individual privacy and permit accurate analysis. They grouped the individuals into classes and cover the relationship between nodes in anonymid graph.

(1) Initialize C to be the set of competitor solutions

(2) Initialize a set S = the unfilled set (the set is to be the ideal arrangement we are constructing).

(3) While C≠and S is (still) not an answer do
(3.1) select x from set C utilizing a voracious strategy

(3.2) erase x from C

(3.3) if {x} S is an attainable arrangement, then

S = S{x} (i.e., add x to set S)
(4) if S is a solution then return S

(5) else return failure.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

VII. Implementation issues:

Implementation is the stage of the project where the theory is accepted and initiation of process to convert theory into working system. Hence it can be considered to be the most important stage in achieving a successful new system and in giving the user, so that the new system will work and be effective.The implementation step involves cautious planning, analysis of the current system and it's constraints on implementation, we can achieve barter and evaluation of changeover methods by using designing methods.

1.Anomalies substructure detection:

Here the objective is to examine the entire graph and report unusual substructures contained it. This technique is for detecting specific, unusual substructure anywhere in the graph.

2.Anomalous sub graph detection:

The subdue method is used for anomalous sub graph detection, where background knowledge is necessary. Subdue can be set to run multiple iterations on a single graph. After each iterations the graph is compressed using discovered substructure. It can say also, every instance of the substructure is replaced by a single vertex. Subdue halts when there is no substructure with more than one instance.

Subdue is an algorithm for detecting repetitive patterns (substructures) within groups.The k-anonymity model prevents the re-identification of nodes and prevents the structural attacks. L-diversity model prevents the attacks (Homogeneity attack and background knowledge attack) and adding of noise nodes to change the degree of the nodes that was done through l-distinct labels of the specified connecting pairs of nodes.

3.User module:

In this module, Users are having verification and security to get to the subtle element which is displayed in the cosmology framework. Before getting to or seeking the points of interest client ought to have the record in that else they ought to enlist first.

4.Verification loss:

We intend to keep data misfortune low. In- arrangement misfortune for this situation contains both structure data misfortune and mark data misfortune. There are some non touchy information's are Loss because of Privacy making so we can't convey full data to the general population.

VIII. Experiment & Result:

Fig: user registration page.

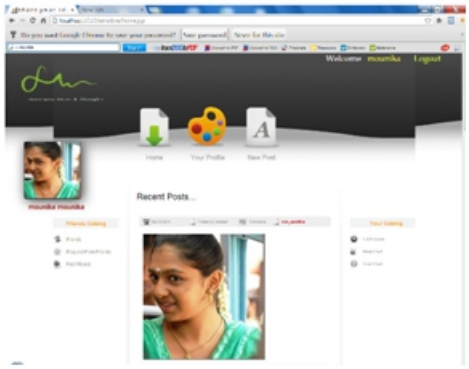

Fig: design tutorials

Fig: put a request to others.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

VIII. Conclusion and future work:

In this paper, a detailed survey is carried out in the anonymization techniques for preserving the individual data in social network. Even though there are more privacy models for preserving the privacy of social network data are developed but the research in this area is still an open issue. The anonymization techniques used to protect the private data of individuals using k-anonymity, l-diversity, t-closeness, KDLD model are adding some of noise nodes to the published graph and make the edge editing technique to implement. The recursive(c,l)-diversity model is modelled to preserve the anonymized data by assigning the sensitive labels to the noise nodes to confuse the attackers and hackers. As the social network data are more complicated than the relational data, the preservation of privacy in social network data is much more challenging task in recent trends. So the critical risks should be carried out for the privacy preserving for relational data and social network data as well.

IX. REFERENCES:

[1] L. Backstrom, C. Dwork, and J.M. Kleinberg, "Wherefore Art Thou r3579x?:Anonymized Social Networks, Hidden Patterns, and Structural Steganography," Proc. Int'l Conf. World Wide Web (WWW), pp. 181-190, 2007.

[2] A.-L. Baraba´si and R. Albert, "Emergence of Scaling in Random Networks," Science, vol. 286, pp. 509-512, 1999.

[3] S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava, "Class-Based Graph Anonymization for Social Network Data," Proc. VLDB Endowment, vol. 2, pp. 766-777, 2009.

[4] A. Campan and T.M. Truta, "A Clustering Approach for Data and Structural Anonymity in Social Networks," Proc. Second ACM SIGKDD Int'l Workshop Privacy, Security, and Trust in KDD (PinKDD '08), 2008.

[5] A. Campan, T.M. Truta, and N. Cooper, "P-Sensitive K-Anonymity with Generalization Constraints," Trans. Data Privacy, vol. 2, pp. 65-89, 2010.

[6] J. Cheng, A.W.-c. Fu, and J. Liu, "K-Isomorphism: Privacy Preserving Network Publication against Structural Attacks," Proc. Int'l Conf. Management of Data, pp. 459-470, 2010.

[7] G. Cormode, D. Srivastava, T. Yu, and Q. Zhang, "Anonymizing Bipartite Graph Data Using Safe Groupings," Proc. VLDB Endowment, vol. 1, pp. 833-844, 2008.

[8] S. Das, O. Egecioglu, and A.E. Abbadi, "Privacy Preserving in Weighted Social Network," Proc. Int'l Conf. Data Eng. (ICDE '10), pp. 904-907, 2010.

[9] W. Eberle and L. Holder, "Discovering Structural Anomalies in Graph-Based Data," Proc. IEEE Seventh Int'l Conf. Data Mining Workshops (ICDM '07), pp. 393-398, 2007.

[10] K.B. Frikken and P. Golle, "Private Social Network Analysis: How to Assemble Pieces of a Graph Privately," Proc. Fifth ACM Workshop Privacy in Electronic Soc. (WPES '06), pp. 89-98, 2006.

[11] S.R. Ganta, S. Kasiviswanathan, and A. Smith, "Composition Attacks and Auxiliary Information in Data Privacy," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 265- 273, 2008.

[12] 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.

[13] G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "A Framework for Efficient Data AnonymizationUnder Privacy and Accuracy Constraints," ACM Trans. Database Systems, vol. 34, pp. 9:1-9:47, July 2009.

[14] J. Han, Data Mining: Concepts and Techniques. Morgan KaufmannPublishers, Inc., 2005.

[15] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis, "Resisting Structural Re-Identification in Anonymized Social Networks," Proc. VLDB Endowment, vol. 1, pp. 102-114, 2008.

[16] E.M. Knorr, R.T. Ng, and V. Tucakov, "Distance-Based Outliers:Algorithms and Applications," The VLDB J., vol. 8, pp. 237-253, Feb. 2000.

[17] N. Li and T. Li, "T-Closeness: Privacy Beyond K-Anonymity andL-Diversity," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE '07), pp. 106-115, 2007.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[18] K. Liu and E. Terzi, "Towards Identity Anonymization onGraphs," SIGMOD '08: Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 93-106, 2008.

[19] L. Liu, J. Wang, J. Liu, and J. Zhang, "Privacy Preserving in SocialNetworks against Sensitive Edge Disclosure," Technical Report CMIDA-HiPSCCS 006-08, 2008.

[20] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L Diversity: Privacy Beyond K-Anonymity," ACM Trans. Knowledge Discovery Data, vol. 1, article 3, Mar. 2007.

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

[22] C.C. Noble and D.J. Cook, "Graph-Based Anomaly Detection,"Proc. Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '03), pp. 631-636, 2003.

[23] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PagerankCitation Ranking: Bringing Order to the Web," Proc. World Wide Web Conf. Series, 1998.

[24] K.P. Puttaswamy, A. Sala, and B.Y. Zhao, "Starclique: GuaranteeingUser Privacy in Social Networks Against Intersection Attacks," Proc. Fifth Int'l Conf. Emerging Networking Experiments and Technologies (CoNEXT '09), pp. 157-168, 2009.

[25] N. Shrivastava, A. Majumder, and R. Rastogi, "Mining (Social) Network Graphs to Detect Random Link Attacks," Proc. IEEE 24thInt'l Conf. Data Eng. (ICDE '08), pp. 486-495, 2008.