Volume 4, Issue 5, May 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Enhanced Security Approach for Centralized and Distributed Social Network by Sequential Clustering

Ms. Sonali M. Khairnar
Department Of Computer Engineering
JSPM's BSIOTR (W), Wagholi
Pune, India

Prof. Sanchika Bajpai
Department Of Computer Engineering
JSPM's BSIOTR (W), Wagholi
Pune, India

Abstract: ***As we know the social media has grown-up rapidly in the past few years. The different social media sites Facebook, Twitter, LinkedIn, and several holds public and confidential information about their users. In order to provide security to the users' social network graphs are anonymized before being published to a third party for data mining or statistical study of privacy preservation of social networks. The aim is to find the anonymized users make the cluster of each and reach at an anonymized view of the network without any of the data holders' information near links among nodes that are organized by other data holders'. For that our work start with the centralized setting which finds the centroid and offer two variants of an anonymization algorithm which is based on sequential clustering and the incremental clustering. Our algorithms considerably beat the SaNGreeA algorithm which is implemented using sequential clustering that is the important algorithm for completing privacy in networks by using clustering. We then devise secure distributed forms of the algorithms. As the best of our knowledge, this one is the first study of privacy preservation in distributed social networks.***

Keyword: Social Networking, Sequential Clustering, Incremental clustering, Privacy Preservation of Data Mining etc

I. INTRODUCTION

The network is nothing but the structure that defines a set of entities and the relations between the different users. A social network that provides the information on entities that is some people and the links between them, which may be relations of friendship, association, correspondence and so on. An information network is the another example that may refer to scientific publications and their reference links. In their maximum basic method, networks are displayed by the different graphs, where the nodes of the graph correspond to the entities and the edges represent relations between them. The social networks may be more difficult than the network. For example, in the network where it describes the interface is asymmetric for example a financial transaction network, the graph would be directed; if the interaction includes more than two parties for example a social network that describes co membership in social clubs then the network would be modeled as a hyper graph in case where there are several types of interaction the edges would be labeled or the nodes in the graph could be attended by attributes that provide demographic information such as age, gender, location, or occupation which could enrich and shed light on the structure of the network. However, the data in such social networks cannot be released as is, since it might contain sensitive information. Therefore, it is needed to anonymize the data prior to its publication in order to address the need to respect the privacy of the individuals whose sensitive information is included in the data.

II. RELATED WORK

Existing System:

In the existing system we cannot maintain privacy to secure the data. The data in such social networks cannot be on the loose as is, since it might contain sensitive and private information. Existing system uses sequential clustering to move the players. A native anonymization of the network, in the sense of removing identifying attributes like names or social security numbers from the data, is insufficient.

Proposed System:

In this project our main aim is to present the extended method for Anonymization of Centralized and Distributed Social Networks by incremental Clustering with improved reliability and performance:

- To present literature review different methods of privacy in anonymizing, present new framework and methods.
- To present the practical simulation of proposed algorithms and evaluate its performances.
- To present the comparative analysis of existing and proposed algorithms in order to claim the efficiency.

We presented incremental clustering algorithms for anonymizing social networks. Those algorithms produce anonymizations by means of clustering with better utility than those achieved by existing algorithms. We devised a secure distributed version

of our algorithms for the case in which the network data is split between several players. We focused on the scenario in which the interacting players know the identity of all nodes in the network, but need to protect the structural information (edges) of the network.

In distributed scenario, each of the players needs to protect the identity of the nodes under his control from the other players. Hence, it is more difficult than Scenario in two manners: it requires a secure computation of the descriptive information loss; and the players must hide from other players the allocation of their nodes to clusters. So we use incremental clustering algorithm for as in network is spread very vastly as to adapt all nodes.

III. SYSTEM IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

Investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

System Architecture

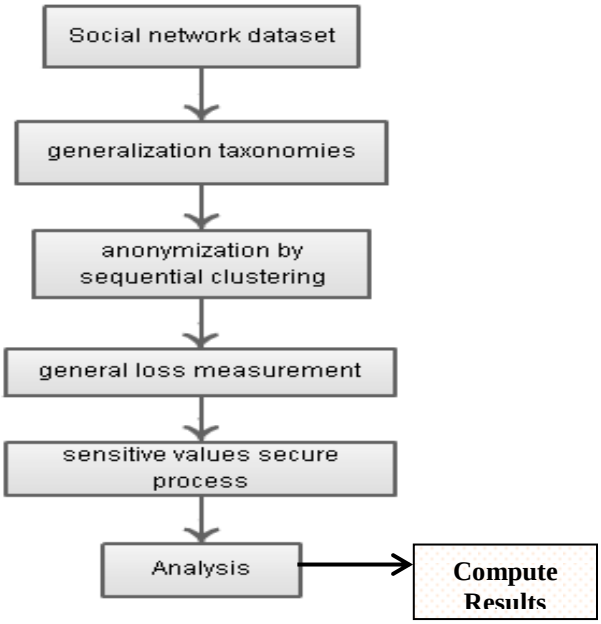

Fig: System Architecture

Main Modules:-

1. Anonymization by clustering:

The study of anonymizing social networks has concentrated so far on centralized networks, i.e., networks that are held by one data holder. However, in some settings, the network data is split between several data holders, or players. For example, the data in a network of email accounts where two nodes are connected if the number of email messages that they exchanged was greater than some given threshold, might be split between several email service providers. As another ex-ample, consider a transaction network where an edge denotes a financial transaction between two individuals; such a network would be split between several banks. In such settings, each player controls some of the nodes (his clients) and he knows only the edges that are adjacent to the nodes under his control.

2. Measuring the loss of information:

Given a social network SN and a clustering C of its nodes, the information loss associated with replacing SN by the corresponding clustered network, SNC, is defined as a weighted sum of two metrics,

$I(C) = w \bullet ID(C) + (1 - w) \bullet IS(C);$

here, $w \in [0,1]$ is some weighting parameter, ID(C) is the descriptive information loss that is caused by generalizing the exact quasi-identifier records R to R, while IS(C) is the structural information loss that is used by collapsing all nodes of V in a given cluster of VC to one super-node.

3. Anonymization by sequential clustering:

we consider anonymizations of a given social network by means of clustering. Let C = {C1,...,CT} be a partition of V into disjoint subsets, or clusters; i.e., $V = \cup T\ t=1Ct$ and $Ct \cap Cs = \emptyset$ for all $1 \leq t \neq s \leq T$. The corresponding clustered graph GC= (VC,EC) is the graph in which the set of nodes is VC= C, and an edge connects Ct and Csin ECiff E contains an edge from a node in Ctto a node in Cs. Each node $Ct \in VC$is accompanied by two pieces of information — |Ct| (the number of original V -

nodes that Ctcontains), and et, which is the number of edges in E that connect nodes within Ct. In addition, each edge {Ct, Cs} ∈ EC is labeled by a weight et,sthat stands for the number of edges in E that connect a node in Ct to a node in Cs.

IV. ALGORITHMS

1. Algorithm

Input: a SN Social network, an integer k

Output: A clustering of SN into clusters of size≥k

Process:

Step 1. Choose a random partition C={C_1,…,C_T} of V into $T := \lfloor N / k_0 \rfloor$

Step 2. $For\ n = 1, \dots, N\ do$

Step 3. Let C_t be the cluster to which v_n currently belongs.

Step 4. For each of the other clusters, C_s, s≠t, compute difference in the information loss, $\Delta_{n:t\to s}$ if v_n would move from C_t to C_s.

Step 5.Let C_{s0} be the cluster for which $\Delta_{n:t\to s}$,is minimal.

Step 6. If $C_t$is a singleton, move v_n from C_t to C_{s0} and remove cluster C_t

Step 7.Else, if $\Delta_{n:t\to s0}$<0, movev_n from C_t to C_{s0}

Step 8.If there exist clusters of size greater than k_1 split each them randomly into two equally-sized clusters.

Step 9.If at least one node was moved during the last loop, go to Step 2 to 8.

Step 10.while there exist cluster of size smaller than k, select one of them and unify it with the cluster which is closest.

Step 11.resulting cluster as output

V. CONCLUSION

Our methods resort to efficient approximation algorithms based on sampling. By sampling we avoid visiting all nodes in the vicinity of a user and thus attain improved performance. The utility of our approach was demonstrated by running experiments on real and synthetic data sets. Further, we showed that our algorithms are able to efficiently estimate the ordering of a list of items that lie on nodes in a user's network providing support to ranking algorithms and strategies. Our research suggests methods for quickly collecting information from the neighborhood of a user in a dynamic social network when knowledge of its structure is limited or not available. Our methods resort to efficient approximation algorithms based on sampling.

ACKNOWLEDGEMENT

I wish to express my sincere thanks and deep gratitude towards my guide Prof. Sanchika Bajpai for her guidance, valuable suggestions and constant encouragement in all phases. I am highly 4 indebted to her help in solving my difficulties which came across whole Paper work. Finally I extend my sincere thanks to respected Head of the department Prof. G.M.Bhandari and Prof. A.C. Lomte [P.G Co-Ordinator] and all the staff members for their kind support and encouragement for this paper. Last but not the least, I wish to thank my Mother for her unconditional love and support.

REFERENCES

[1] A. Arenas, L. Danon, A. Díaz-Guilera, P. M. Gleiser, and R. Guimerá, "Community Analysis in Social Networks," *The European Physical Journal B*, Vol. 38, Number 2, pp. 373-380, 2004.

[2] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography," in *Proc. WWW'07*, pp. 181-190, 2007.

[3] S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava, "Classbased Graph Anonymization for Social Network Data," in *Proc. VLDB'09*, pp. 766-777, 2009.

[4] A. Campan and T.M. Truta, "Data and Structural k-Anonymity in Social Networks," Proc. Second ACM SIGKDD Int'l Workshop Privacy, Security, and Trust in KDD (PinKDD), pp. 33-54, 2008.

[5] J. Goldberger and T. Tassa, "Efficient Anonymizations with Enhanced Utility," Trans. Data Privacy, vol. 3, pp. 149-175, 2010.

[6] S. Hanhija¨rvi, G. Garriga, and K. Puolamaki, "Randomization Techniques for Graphs," Proc. Ninth SIAM Int'l Conf. Data Mining (SDM), pp. 780-791, 2009.

