International Journal of Information Technology and Management
Vol. 12, Issue No. 1, February-2017, ISSN 2249-4510

Use of Data Mining Method for Secure Privacy in Social Networking Sites

Kale Deepali Anil[1]* Dr. Suneel Kumar[2]

[1]Ph.D. Research Student, MUIT, Lucknow

[2]Research Guide, MUIT, Lucknow

Abstract – Development of online social networks and publication of social network data has led to the risk of leakage of confidential information of individuals. This requires the preservation of privacy before such network data is published by service providers. Privacy in online social networks data has been of utmost concern in recent years. Hence, the research in this field is still in its early years. Several published academic studies have proposed solutions for providing privacy of tabular micro-data. But those techniques cannot be straight forwardly applied to social network data as social network is a complex graphical structure of vertices and edges. Social network has gained remarkable attention in the last decade. Accessing social network sites such as Twitter, Facebook LinkedIn and Google+ through the internet and the web 2.0 technologies has become more affordable. People are becoming more interested in and relying on social network for information, news and opinion of other users on diverse subject matters. Data mining techniques are used for information retrieval, statistical modelling and machine learning. These techniques employ data pre-processing, data analysis, and data interpretation processes in the course of data analysis. This survey discusses different data mining techniques used in mining diverse aspects of the social network over decades going from the historical techniques to the up-to-date models, including our novel technique.

Keywords: Data Mining, Secure, Privacy, Social Networking Sites, Technique, etc.

----------X----------

INTRODUCTION

Social network is a term used to describe web-based services that allow individuals to create a public/semi-public profile within a domain such that they can communicatively connect with other users within the network. Social network has improved on the concept and technology of Web 2.0, by enabling the formation and exchange of User-Generated Content. Simply put, social network is a graph consisting of nodes and links used to represent social relations on social network sites. The nodes include entities and the relationships between them forms the links (as presented in Fig. 1).

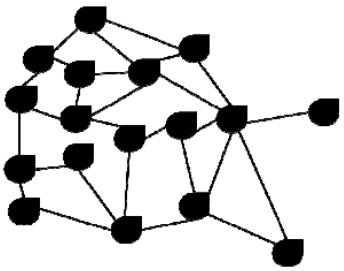

Fig. 1- Social Network nodes

Social networks are important sources of online interactions and contents sharing subjectivity assessments, approaches, evaluation, influences, observations, feelings, opinions and sentiments expressions borne out in text, reviews, blogs, discussions, news, remarks, reactions, or some other documents. Before the advent of social network, the homepages was popularly used in the late 1990s which made it possible for average internet users to share information (Xintao, et. al., 2010). However, the activities on social network in recent times seem to have transformed the World Wide Web (www) into its intended original creation. Social network platforms enable rapid information exchange between users regardless of the location. Many organisations, individuals and even government of countries now follow the activities on social network (Christopher, 2011). (Zheleva and Getoor, 2011). (Fire, et. al., 2012). The network enables big organisations, celebrities, government official and government bodies to obtain knowledge on how their audience reacts to postings that concerns them out of the enormous data generated on social network (Masoumzadeh and Joshi, 2012).

REVIEW OF LITERATURE:

During the last decade social network have become not only popular but also affordable and universally-acclaimed communication means that has thrived in making the world a global village. Social network sites are commonly known for information dissemination, personal activities posting, product reviews, online pictures sharing, professional profiling, advertisements and opinion/sentiment expression. News alerts, breaking news, political debates and government policy are also posted and analysed on social network sites. It is observed that more people are becoming interested in and relying on the social network for information in real time (Tassa and Cohen, 2013). Users sometimes make decisions based on information posted by unfamiliar individuals on social network increasing the degree of reliance on the credibility of these sites. Social network has succeeded in transforming the way different entities source and retrieve valuable information irrespective of their location. Social network has also given users the privilege to give opinions with very little or no restriction.

Categories of Privacy Breach: The privacy breaches in social networks can be categorized into three types: i. Identity disclosure - Identity disclosure occurs when an individual behind a record is exposed. This type of breach leads to the revelation of information of a user and relationship he/she shares with other individuals in the network. ii. Sensitive link disclosure - Sensitive link disclosure occurs when the associations between two individuals are revealed. Social activities generate this type of information when social media services are utilized by users. iii. Sensitive attribute disclosure – Sensitive attribute disclosure takes place when an attacker obtains the information of a sensitive and confidential user attribute. Sensitive attributes may be linked with an entity and link relationship.

Challenges in Preserving Privacy in Social Network Data Publishing ensuring: privacy for social network data is difficult than the tabular micro-data because: a) Modeling of background knowledge of adversaries is difficult in social network data than tabular micro-data. In tabular micro-data, users are identified by linking quasiidentifiers from whereas in social network information from various sources such as labels of vertices and edges, subgraphs, and neighborhood graphs can be used to identify individuals. b) Information loss is the metric which measures the amount of distortion. In tabular micro-data information loss can be measured using the sum of information loss in individual records. Since, a social network is a graphical structure with a set of vertices and edges hence it is difficult to compare two social networks by comparing the vertices and edges individually. Anonymized social network and original social networks which have the same number of vertices and edges may have very different properties like betweenness, connectivity, and diameter (Heatherly, et. al., 2013). Information loss and anonymization quality can be measured in different ways. c) Development of privacy preserving techniques in social network data is difficult than for relational data. Tabular micro-data is anonymized using divide-and-conquer techniques whereas social network is a structure of nodes and edges, any changes in labels or edges may have an effect on the neighborhoods of other vertices and edges.

The methods proposed for tabular micro-data cannot be directly applied to social network data due to the connectivity between vertices in the graph network as compared to independent nodes in tabular data. In micro-data, each tuple is independent, but the vertices and edges in a social network are linked to each other. An adversary can use the information regarding network structure to violate the privacy of users. So there is a need is to develop a technique that can ensure the privacy of the entities in social network data publishing.

Privacy Preserving Techniques – Micro Data: Significant work has been done for preserving privacy in tabular micro-data. Models like K-anonymity, Diversity, T-closeness have been proposed which have shown good results in anonymization. Fig. 2 briefs the three models, their properties and drawbacks.

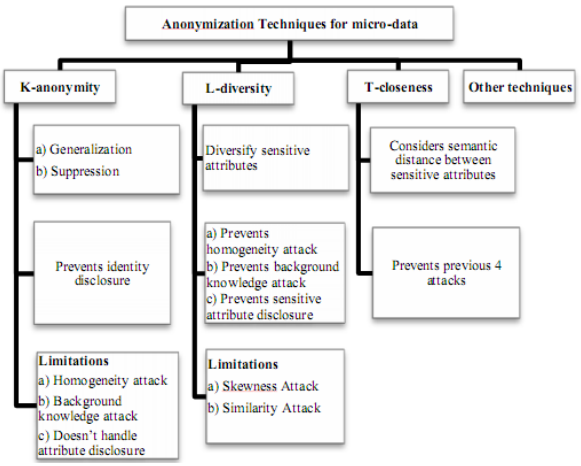

Fig. 2: Existing Privacy Preserving Techniques for Tabular micro-data

Social Network – Power to the Users: Social sites have undoubtedly bestowed unimaginable privilege on their users to access readily available never-ending uncensored information. Twitter, for example, permits its users to post events in real time way ahead the broadcast of such events on traditional news media. Also, social network allow users to express their views, be it positive or negative (Cheng and Sandhu, 2013). Organizations are now conscious of the significance of consumers' opinions posted on social network sites to the patronage of their products or services and the overall success of their organisations. On the other hand, important

personalities such as celebrities and government officials are being conscious of how they are perceived on social network. These entities follow the activities on social network to keep abreast with how their audience reacts to issues that concerns them. Considering the enormous volume of data being generated on social network, it is imperative to find a computational means of filtering, categorizing, classifying and analysing the social network contents.

Recommender System in Social Network Community: Based on the mutuality between nodes in social network groups, collaborative filtering (CF) technique, which forms one of the three classes of the recommender system (RS), can be used to exploit the association among users. Items can be recommended to a user based on the rating of his mutual connection. Where CF's main downside is that of data sparsity, content-based (another RS method) explore the structures of the data to produce recommendations. However, the hybrid approaches usually suggest recommendations by combining CF and content-based recommendations (Sun, et. al., 2010). The experiment in proposed a hybrid approach named Entrée C, a system that pools knowledge-based RS and CF to recommend restaurants. The work in improved on CF algorithm by using a greedy implementation of hierarchical agglomerative clustering to suggest forthcoming conferences or journals in which researchers (especially in computer science) can submit their work.

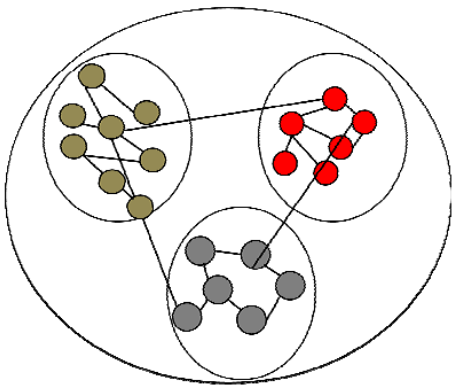

Fig. 3- Social Network Community Structure

Semantic Web of Social Network: The Semantic Web platform makes knowledge sharing and re-use possible over different applications and community edges. Discovering the evolvement of Semantic Web (SW) enhances the knowledge of the prominence of Semantic Web Community and envisages the synthesis of the Semantic Web (Beach, et. al., 2010). The work in employed Friend of a Friend (FOAF) to explore how local and global community level groups develop and evolve in large-scale social networks on the Semantic Web. The study revealed the evolution outlines of social structures and forecasts future drift. Likewise application model of Semantic Web-based Social Network Analysis Model creates the ontological field library of social network analysis combined with the conventional outline of the semantic web to attain intelligent retrieval of the Web services (Yan, et. al., 2010). Furthermore, Voyeur Server improved on the open-source Web-Harvest framework for the collection of online social network data in order to study structures of trust enhancement and of online scientific association. Semantic Web is a relatively new area in social network analysis and research in the field is still evolving.

CONCLUSION:

It became evident from the literature that privacy of users is the main concern and topic of research now days. Various models proposed for tabular micro-data have been adopted for preserving privacy of social network data. Techniques like Anonymity, L-diversity, and integrated K-anonymity L-diversity have been used till now but these techniques lead to substantial information loss. So, there is a scope of improvement of the techniques that provide privacy preservation with minimum information loss and better utility of released data. Different data mining techniques have been used in social network analysis as covered in this survey. The techniques range from unsupervised to semi-supervised and supervised learning methods. So far different levels of successes have being achieved either with solitary or combined techniques. The outcome of the experiments conducted on social network analysis is believed to have shed more light on the structure and activities of social network. The diverse experimental results have also confirmed the relevance of data mining techniques in retrieving valuable information and contents from huge data generated on social network. Future survey will tend to investigate novel state-of-the-art data mining techniques for social network analysis.

REFERENCES:

Aaron Beach, Mike Gartrell, Richard Han (2010). "q-Anon: Rethinking Anonymity for Social Networks", In Proc. of IEEE Second International Conference on Social Computing (SocialCom), Minneapolis, MN, pp. 185 – 192.

Amirreza Masoumzadeh, James Joshi (2012). "Preserving Structural Properties in Edge-Perturbing Anonymization Techniques for Social Networks", In: IEEE Transactions On Dependable And Secure Computing, Vol. 9, No. 6, pp. 877-889.

Christopher C. Yang (2011). "Preserving privacy in social network integration with τ-tolerance", In Proc. of IEEE International Conference on

Intelligence and Security Informatics (ISI), Beijing, pp. 198 – 200.

Elena Zheleva, Lise Getoor (2011). “Privacy in Social Networks: A Survey”, In: Social Network Data Analytics, Springer US, pp. 277-306.

Jinyuan Sun, Xiaoyan Zhu, and Yuguang Fang (2010). “A Privacy-Preserving Scheme for Online Social Networks with Efficient Revocation”, In Proc. of INFOCOM, IEEE, San Diego, CA, pp. 1-9.

Michael Fire, Dima Kagan, Aviad Elishar, and Yuval Elovici (2012). "Social Privacy Protector - Protecting Users’ Privacy in Social Networks,", In Proc. of the Second International Conference on Social Eco-Informatics (SOTICS), Venice, Italy.

Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham (2013). “Preventing Private Information Inference Attacks on Social Networks”, In: IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 8, pp. 1849-1862.

Tamir Tassa, Dror J. Cohen (2013). “Anonymization of Centralized and Distributed Social Networks by Sequential Clustering”, In: IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 2, pp. 311- 324.

Xintao Wu, Xiaowei Ying, Kun Liu, Lei Chen (2010). “A Survey of Privacy-Preservation of Graphs and Social Networks”, In : Managing and Mining Graph Data, Advances in Database Systems, Vol. 40, pp. 421-453.

Yan Zhu, Zexing Hu, Huaixi Wang, Hongxin Hu, GailJoon Ahn (2010). “A Collaborative Framework: for Privacy Protection in Online Social Networks”, In Proc. of 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), Chicago, IL, pp. 1 – 10.

Yuan Cheng, Ravi Sandhu (2013). “Preserving User Privacy from Third-party Applications in Online Social Networks, In Proc. of 22nd international conference on World Wide Web companion, Geneva, Switzerland, pp. 723-728, 2013.

Corresponding Author

Kale Deepali Anil*

Ph.D. Research Student, MUIT, Lucknow

E-Mail – d_a_kale@yahoo.co.in