COMMUNITY PRIVACY PRESERVATION IN DYNAMIC SOCIAL NETWORKS

[1] Neelima Kokkiligadda, [2] Prof. Valli Kumari Vatsavayi
[1,2] Dept. Computer Science and System Engineering , Andhra University College of Engineering
Visakhapatnam, India
[1] neelima561@gmail.com

***Abstract*— Facebook, Twitter, LinkedIn, Instagram are well known social networking sites. These sites gather data regarding their user's interests, disinterests, location, profession etc. This data may contain sensitive information about the user. This data is used for research and business purposes. So this social network data should be made anonymize before it is made available to the third parties. This paper deals with community preservation in dynamic social networks and for this it considers social network at two different time periods. Fast k-degree anonymity algorithm is used to anonymize the initial social network and to produce the anonymized social network. To form communities from the initial and anonymized social networks Louvain community detection method is used. Then the community preservation is performed between communities from initial and anonymized graphs. And the percentage community preservation is obtained for different time periods.**

***Keywords*— Privacy, Preservation, community, detection, Anonymization, social network.**

I. INTRODUCTION

Facebook, Twitter, LinkedIn, Instagram are well known social networking sites. These sites have huge number of users and theses sites allow users to share ideas, pictures, posts, activities, events, and interests with people in their network. And the data shared by the users on these social networks is very valuable and Social network sites started using this data for business purposes. At the same time this social network data contains sensitive information related to corresponding individual user. Using this data for business purpose, increasing the risk of individual identity disclosure. So this social network data should be made anonymize before it is used for the research and marketing purposes. The previous works [1] related to this area focuses only on anonymity models in static social networks.

This paper focuses on community privacy preservation in dynamic social networks using community detection algorithm and anonymity model. The initial social network graph which represents set of nodes and the edges is anonymized using the Fast k-degree anonymity algorithm. As in real time the social network graph is not constant and it structure changes according to the time. The initial social network graph is anonymized at two different time periods. To extract communities from initial and anonymized social network graph Louvain community detection algorithm considered. Community privacy preservation is performed between communities from initial and anonymized social networks at different time periods.

II. ANONYMIY MODEL: FAST K – DEGREE ANONYMITY

In this paper we are using Fast k-Degree Anonymity model [2]. This anonymity model makes the graph as anonymized graph by adding edges in a greedy fashion. First it takes the graph nodes in descending order of their degrees. It forms the vertices into clusters according to the k-value.

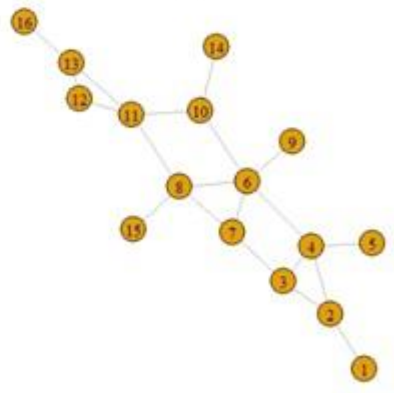

Fig 1. Initial Social Network

After forming the separate groups it simply anonymizes the each group by adding the edges between nodes within the group. After this process it checks for the unanoymized groups and they are made anonymized by adding the edges between vertices within among those groups. At the end of this process all groups are anonymized. In this way Fast K-Degree anonymity model anonymizes a graph. Initial and fast k-degree anonymized graphs are shown in Fig 1 and Fig 2.

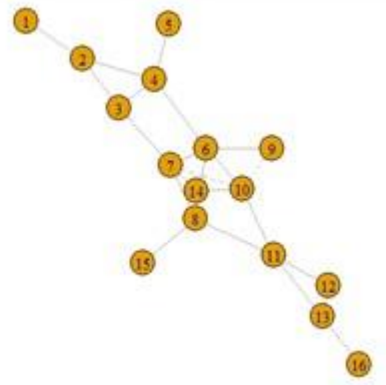

Fig 2. Anonymized Social Network

III. COMMUNITY DETECTION

A social network graph consists of vertices and edges as a normal graph. Here the vertices are called nodes and each node represents one user and the edges represent the relation between the nodes. Generally these are formed into groups which are namely called communities. Communities are groups of nodes from social network likely have similar properties or characteristics. Community detection is a process of extracting communities from social networks. In this paper we are using Louvain community detection [3].

A. Louvain community detection

The main idea behind this community detection is the optimization of modularity [4] [5]. Modularity is a quality function [6] that can be computed for a graph partitioned in communities. The value of modularity lies between -1 and 1 that measures the density of edges inside communities to edges outside communities.

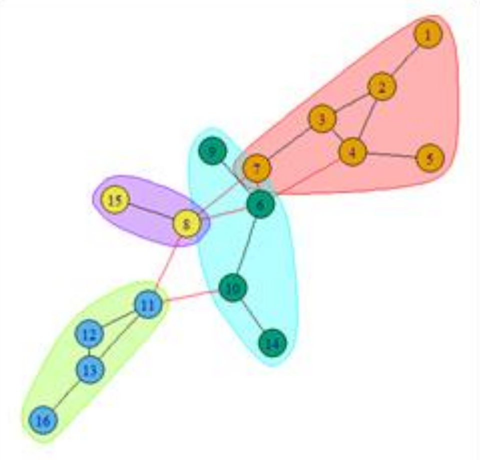

Fig 3. Initial Network Communities

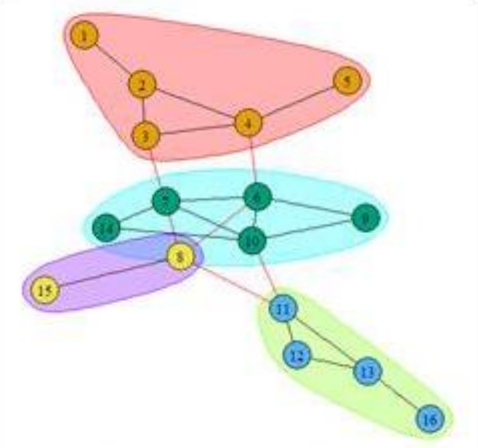

Fig 4. Anonymized Network Communities

In this method each node is assigned to one community. Then the modularity gain of each community is maximized by moving nodes between those communities. This step is stopped when there is no change in modularity gain with the movement of nodes. After this process the network obtained from the first step is used and a weighted network is created. In this weighted network, one node represents a community from the original network, and weights are added to edges to represent the number of original edges that are collapsed into a super edge. After the completion of this step again the first step is implemented. This process repeated iteratively until the modularity gain is maximized. Communities obtained by Louvain method from initial and anonymized graphs are shown in fig 3 and fig 4.

IV. COMMUNITY PRESERVATION

In community detection communities are extracted by using community detection algorithm from initial and anonymized social networks. These communities are needed to be compared to check whether the communities are same after and before anonymization. For this purpose community preservation is used. Community preservation is a technique that is used to verify the similarity between the communities from initial and anonymized social networks. In this paper privacy preservation is performed by using privacy preservation at community level (PPCL).

A. Privacy preservation at community level (PPCL)

In this method the communities extracted from the initial and anonymized social networks are compared with each other to calculate percentage privacy preservation. The percentage privacy preservation is calculated by dividing the common nodes present in both the communities with the number of nodes present in the initial communities. For communities [1,2,3,4,5,7] and [1,2,3,4,5] PPCL is calculated as 5/6 = 83.33%. Where 5 is number of nodes common between communities from initial and anonymized social networks and 6 is number of nodes of a community from initial network. Percentage preservation is calculated for each community and then average percentage preservation is calculated from all individual community preservation. The average values are represents percentage preservation value for a social network.

The Percentage preservation values obtained by PPCl for initial and anonymized networks are given in table 1, 2, 3. Their percentage preservation values are 95.83%, 89.58, 74.15. These values represent percentage preservation of initial and anonymized social networks at time t1, t2 and t1t2. Here t1 represents a time period 1 at which percentage preservation values obtained from initial and anonymized social networks. Here t2 represents a time period2 at which percentage preservation values obtained from initial and anonymized social networks. Here t1t2 represents a time period 1 and 2 at which percentage preservation values obtained anonymized social network at time t1 and anonymized social network at time t2. Here comi represents communities from theinitial social networks and comi’ represents communities from the anonymized social network.

TABLE 1. Privacy Preservation at Community level at time t1

Community	com$_i$	com$_i'$	*PPCL*
1	[1,2,3,4,5,7]	[1,2,3,4,5]	83.33%
2	[11,12,13,16]	[11,12,13,16]	100%
3	[6,9,10,14]	[6,7,9,10,14]	100%
4	[8,15]	[8,15]	100%

TABLE 2. Privacy Preservation at Community level at time t2

Community	com$_i$	com$_i'$	*PPCL*
1	[1,2,3,4,5,7]	[1,2,3,4,5]	83.33%
2	[11,12,13,16,17, 20]	[11,12,13,16,17,20]	100%
3	[6,9,10,14,19]	[6,9,10,14,19]	100%
4	[8,15,18]	[7,8,15,18]	75%

TABLE 3. Privacy Preservation at Community level at time t1t2

Community	com$_i$	com$_i'$	*PPCL*
1	[1,2,3,4,5]	[1,2,3,4,5]	100%
2	[11,12,13,16]	[11,12,13,16,17,20]	66.6%
3	[6,7,9,10,14]	[6,7,9,10,14,19]	80%
4	[8,15]	[7,8,15,18]	50%

V. EXPERIMENTS AND RESULTS

The main purpose of this work is to study the community preservation between the initial and anonymized social networks at different time periods. The community preservation is performed on two different datasets. First dataset represents the college students studying in same college from different cities. Dataset1 is student dataset which contains 500 nodes and 1406 edges. In this dataset each node represents student and the relationship between them is represented by an edge which shows that both students belongs to same city. Second dataset represents the employees from same organization with different streams in graduation. Dataset2 is employee dataset which contains 1130 nodes and 2930 edges. In this dataset each node represents an employee and the relationship between them is represented by an edge which shows that both employees having degree in same stream.

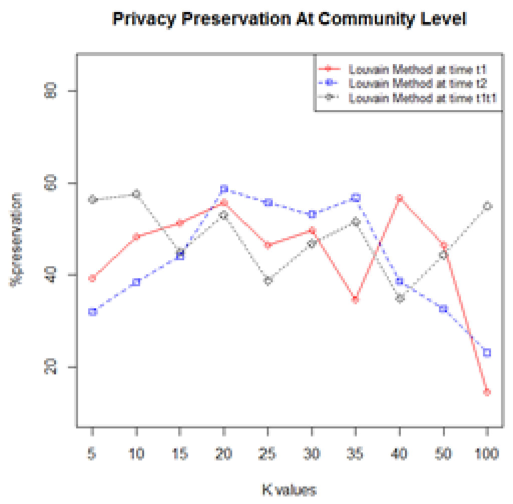

Fig 5. PPCL for Student dataset

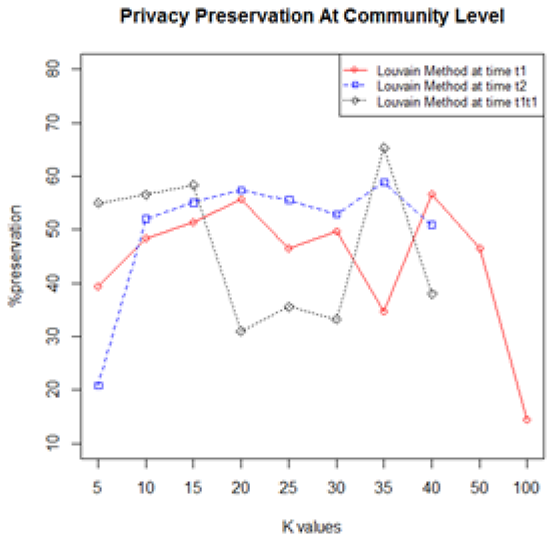

Fig 6. PPCL for Employee dataset

Graphs are drawn for the obtained results by plotting the percentage preservation values on y-axis and k values on x-axis. Here the k values are taken from 5 to 100 and k represents the degree of anonymization. Fig 5 represents the percentage preservation values for student dataset and Fig 6 represents the percentage preservation for employee dataset.

Here the graphs are drawn at two time periods t1 and t2. At time t1 percentage preservation values obtained from initial and anonymized social networks. At time t2 percentage preservation values obtained from initial and anonymized social networks. At time t1t2 percentage preservation values obtained from anonymized social network at time t1 and anonymized social network at time t2.

CONCLUSIONS

In this paper we studied how well communities are preserved in dynamic social networks. Here we considered two different time periods to measure privacy preservation in dynamic social networks. For social network anonymization Fast k-degree anonymization technique is used. Privacy preservation is obtained by using privacy preservation at community level. The obtained percentage preservation value at t1 and t2 are good and these values showing that communities in dynamic social networks are well preserved by the preservation technique.

REFERENCES

[1] A.Campan, Y. Alufaisan, and T. M. Truta. Preserving communities in anonymized social networks. Trans. Data Privacy, 8(1):55–87, Dec. 2015.

[2] Y. B. S. Lu X. Song. Fast Identity Anonymization on Graphs. September 2012.

[3] R. L. E. L. Vincent D. Blondel, Jean-Loup Guillaume. Fast unfolding of communities in large networks. J.Stat. Mech. (2008) P10008, 2.

[4] M. E. J. Newman. The structure of scientific collaboration networks. 98(2):404–409, January 2001.

[5] A. Noack. Modularity clustering is force-directed layout. Phys. Rev. E 79, 026102 (2009), 1:9, July 2008

[6] S. Fortunato. Community detection in graphs. Physics Reports 486, 75-174 (2010), 1