Comparison of K-automorphism and K^2-degree Anonymization for Privacy Preserving in Social Network

Sumit Kumar Chaurasia
School of information technology
Rajiv Gandhi Proudyogiki Vishwavidyalaya
Bhopal India

Nishchol Mishra
School of information technology
Rajiv Gandhi Proudyogiki Vishwavidyalaya
Bhopal India

Sanjeev Sharma
School of information technology
Rajiv Gandhi Proudyogiki Vishwavidyalaya
Bhopal India

ABSTRACT

Social networking sites are extensively useful for communicating with the real world. The conversation among these sites accomplishes an extreme amount of data in the network. These large amount of data contain vulnerable information that is sharing between social users through links or edges. A broad use of social network creates security and privacy issues of a network. Many of social users unguarded about the risks, which caused by extrovert their sensible data, make network bunce for identity, and link disclosure. Simply the privacy of users is preserve by removing the identified element of users but it is not enough for user’s privacy through attacks, which have some prior knowledge about users. This paper mainly concerned with friendship and structural attack on user’s privacy, which disclose user’s identity and link information.. Detailed analysis is done regarding k^2-degree and k-automorphism methods for protecting the privacy from these attacks and the utility such as average shortest path and cluster coefficient were also calculated for these two method.

Keywords

Social network, k2-degree, k-automorphism, Anonymization.

1. INTRODUCTION

Now a day’s Social networking is widely used for communication across the world. In recent year most of the traffic in digital network were increased due to the use of social networking. Social sites like Facebook, Flicker, YouTube, MySpace and many more provide a way of using digital data, which is available online. This data include multimedia as photo sharing, audio and video sharing, uploading and downloading, entertainment as virtual word and online gaming and microbloging as chatting and many others. Social networking sites create online relationship between groups and communities where users share their personal comment on religion, politics. These data sharing platform are represented through graph, where users and communication link is represent by nodes and links, sensitive information of users is represented as label on either node or link. Social network sites easily provide the way of communication as a result, there popularity is rapidly increases. This popularity increment cause the information loss and identity and link disclosure of a users, it also break users privacy and lead to harmful effect like criminal charges. This problem about user’s sensitive information disclosure to intruder has worst to brush of online social network. Identity disclosures happen when an intruder is able to determine that a profile in a social network match with individual real-world entity. [1] Some issues related to identity disclosure are as follows:

a. Mapping query – in mapping query, intruder determines which profile represents a particular user in a set of individual profiles
b. Existence query – in existing query, intruder finds out if a user has already a profile in a social network or not.
c. Co reference query – in co-reference query intruder picks up that the two user profiles on social network refer to the same real-world user or they refer to different real-world user.

Link disclosure [9] is happening when an intruder is determined any sensible relationship among users in the network or a relationship that users want to hide from the real world. This sensible relationship among users may carry various sensitive communication data information such as phone call, messages, chatting, tagging, comments etc. their relationship such as knowing that which group or community a particular user belong and which type of relation they have with the group. If some user has belonged to same group may also be sensitive in nature. Sometimes link disclosure can lead to identity, attribute, and link disclosure. So hiding affiliation is more important to preserve the privacy of users in social networks.
This paper focuses on two type of attack namely friendship attack [3] and structure attack [7] where the intruder know some prior information about user’s and disclose their privacy. These attacks on privacy are removed by two proposed method, one is k^2-degree and another is k-automorphism. In friendship attack intruder get friend list of target person and find out sensible relationship between two individual user’s and re-identify user’s and their friends in publish social network and theirs hobbies, activities and political and religious belief.

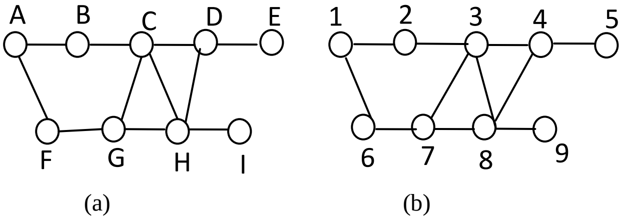

Figure 1: (a) original social network G (b) anonymized social

network G'. For example in figure 1, if intruder find out information about degree of a vertex corresponding to the user, intruder cannot find out any one in the anonymize network but if he know that user b and c are friend then they are uniquely identified by the vertex degree pair. For protecting against friendship attack, concepts of k2-degree anonymity is introduced which uses integer programming and scalable approach to find out a valuable solution.
Whereas in structural attack, identity of user is disclose if intruder find an individual entity with high probability or some structural information. In publish anonymized graph, if sensible link of target user’s are located then this information is disclosed. For example if two users are connecting with the

link and share, sensitive information and intruder find out this relationship to break their privacy. In Breaching privacy through structural attack the intruder uses some background knowledge based attack to re-identify the target user; these are degree attack, sub-graph attack, 1-neighbour-graph attack and hub-fingerprint attack.

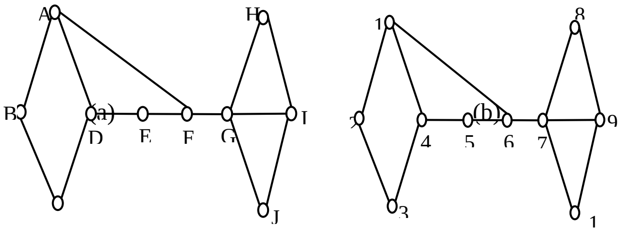

Figure 2: (a) original network (b) anonymize network

For example in figure 2 if a network G, and a anonymize network G' is obtain by removing the individual name in G. If an attacker target on vertex G and if he know that G has four neighbour, he can uniquely identified G is vertex 7 in G'. To protect from this type of attack, k-automorphism method is introduce which show that there is no structural difference between nodes and it's symmetric vertexes and target user can not be identified with a probability higher than 1/k. It uses k-match, graph alignment, edge copy, graphs partitioning and vertex-id generalization approach for privacy protection.

2. RELATED WORK

Recently many algorithms were proposed for protecting identity and link information against various attacks, in this way, Liu. et.al [10], propose k-degree anonymization for preserving from identity disclosure in a social network. In uses, released social networks graph which simply removing the identity of user which protect information disclosure. Author proposed a graph anonymization algorithm in which graph is k degree anonymous if every node v, in graph G there exist at least k-1 node with the same node degree as a v.

In social network, the link re-identification is crucial issue for user's privacy. Link re-identification is generally occur when the sensitive data passes throw edge or link in graph is read by unauthorized users. To protect link re-identification in the network from unauthorized access, the Elena.et.al [11] proposed an anonymization method where the sensitive relationship between users are remove and then released this anonymized graph in the network. Author consider the node data that is anonymized by the k-anonymization [17], and has also test the influence of data attribute on sensitive relationship and analyze different anonymize technique like node anonymization, edge anonymization, intact edge, partial edge removal, cluster edge anonymization and cluster edge anonymization with constraints for privacy preserving. As a result they compare each anonymization method with other and give that if threshold value is keeping higher then all edge in different classes preserve privacy much better than deleting edges and if the value of k is less then privacy preservation is lower.

In a network if unauthorised user is able to determine graph partition and edge link between two individual, then privacy of users get breaches. To protect the privacy of these edges Lijie.et.al [6] proposed edge anonymity, called graph confidence, which define vertex description type (VDT), a topological feature of vertexes. They analyze edge disclosure and obtain those edge disclosures that occur in the graph. In edge anonymity algorithm two basic operation edge swaps and edge deletions are perform. They implemented a vertex k-degree anonymity algorithm [10]: the priority algorithm with the probing scheme using degree as a VDT and edge deletion as anonymization strategy. They also implemented the simulated annealing algorithm [8], which searches for anonymous graph that optimizes a likelihood estimate.

To protect the privacy of link, node and k-security, author proposed the k-isomorphism [5] in which they firstly take graph G = {V, E}, |V | is a multiple of k and adding no more than k − 1 dummy vertices in the graph. This node and link problem definition is based on the notion of k-security and neighbourhood attack graph. This anonymization method depends on adversary knowledge and utility of algorithm is depending on degree of all vertices, path length and cluster coefficient of graph. This anonymization graph is able to preserve the essential information of users

The link information discloses between two users is harmful regarding user privacy if any attacker knows the prior knowledge like degree of user or structure of user. Structural information like edge addition, deletion or edge swapping, which derive the structure of network may also lead the privacy breaches in the network. To protect from link disclosure Author proposes Subgraph wise perturbation and perturbation process [2], in which the given graph is partition into local sub-graphs and graph wise perturbation is apply to each sub graph.

Machanavajjhala. et. al [13], determines privacy breaches because of attribute discloser, these attribute contain either single value or combination of two values, which are also called quasi-identifier and sensitive data. Author performed two simple attacks on a k-anonymized dataset. First, that there is little diversity in sensible attribute so an intruder can breach these sensible attributes because of lack of diversity which is also known as homogeneity attack. In homogeneity attack both the users and intruder are belonging to same group and Secondly, if intruders have any background knowledge, then it causes serious problems, as k-anonymity does not guarantee privacy against intruders using background knowledge. In a background, attack users may belong to same or different affiliation group.

The *l*-diversity is unable to prevent attribute disclosure against skewness and similarity attack. Similarity attack is accruing when attribute value are different but structurally similar in equivalent class and skewness attack is accruing when *l* diversity does not prevent the attribute discloser. To overcome this limitation of *l*-diversity, Li.et.al proposed a method call t-closeness [14]. It is defined as the distance among the distribution of a sensible attribute in any class, and the distribution of the attribute in the whole table is no more than a threshold t. A table has t-closeness if all equivalence classes have t-closeness. In t-closeness primary requirement is the distribution of sensible attribute in any equivalence class and the distance should not greater than t.

Zhou and pai discuss a new attack called neighbourhood attack [12], based on sub-graph in which intruder know that how many neighbour of target person are in 1 hop distance. The intruder re-identifies the target user in the network. This type of attack is resolved by anonymization method in which no user is re-identify by intruder with probability higher than 1/k.

Chil.et.al [4] introduced a new type of attack called community identity of users. Community identity represents the user's personal view about their community. A structural diversity method is introduced to protect from this attack in which at least k community contain vertices with same degree. Another problem is introduced after applying structural diversity method known as k-structural diversity anonymization and integer programming formulation method is introduce to find the optimal solution of k-structural diversity anonymization.

3. PROBLEM FORMULATION

Consider that social network is a simple graph G = (V, E), where V is the set of node or an individual user and the E is the set of edges or the links between the two users. We perform an analysis between two methods namely friendship attack and structure attack.

3.1 Friendship Attack- Friendship attack is based on the vertex degree of vertexes connected through an edge. Intruder on a graph uses degree of vertices connected by the edge to re-identify related user and find out the relationship between them based on the information available on the network. In friendship attack a target user A and the degree pair information $D^2 = (d_1, d_2)$, and friendship attack (D^2, A) exploits D^2 to identify a vertex v_1 corresponding to A in G, such that v_1 connects to another vertex v_2 in G with the degree pair $(d_{v1}, d_{v2}) = (d_1, d_2)$. In social network G, it is easier for the adversary to identify A from the candidate vertices when the number of candidate vertices is small.

3.2 Structural Attack-

In a social network when the network is published after removing the identifier attribute, the target person is still be recognize in the network by having structural information, this lead to identity and link disclosure in structural attack. Attacker launch a query in the given graph network and this launched query match with limited number of vertices in the network, than target person is uniquely identified in the network and this query is based on the structure of the target. In structure, the author uses four type of attack [8, 12, 10].

3.2.1 Degree attack- In degree attack an attacker know the degree of user that means he know how many people are connected with this particular user in the network. For example if any user has a degree 7, that mean this user is connect 7 other people in the network

3.2.2 Sub-graph attack- In this type of attack user know the structure of the user and find the structure in the graph or find out which sub graph match with the user structure in the network.

3.2.3 1-nehabour-graph attack- In this attack attacker knows about the neighbour of target person and then he find the neighbour of the target person to determine the actual target user.

3.2.4 Hub fingerprint attack- In this type of attack if attacker know the distance between the user and the hub then target person is easily discover.

4. METHOD USED

4.1 Protecting from friendship attack

For protecting the network again friendship attack, k^2-degree anonymization [3] is proposed in which vertices re-identification is not more than the probability to 1/k. Apart from this the integer programming for k^2-degree is also proposed and aim to find optimal solution for small network. A graph G is k^2-degree anonymous if every vertex with an incident edge of degree pair (d_1, d_2) in G, there exist at least k – 1 other vertices, such that each of the k − 1 vertices also has an incident edge of the same degree pair. k^2 – degree has following property against the adversary who have information $D^2 = (d_1, d_2)$ and made friendship and vertex degree attack as:

(i). if a anonymize graph is k^2-degree then it is also in the k-degree anonymous

(ii). if a anonymize graph is k_1^2 - degree then it is also k_2^2 - degree anonymize for every $k_2 \leq k_1$. For archiving the k^2-degree anonymity the edge addition and edge deletion operation is performed on the graph and minimizes the cost of the graph that is increased after the addition or deletion operation. Proposed outcome of an anonymized graph has same set of vertexes as in original graph and the anonymization cost G, G' = ω |E'/E| + (1- ω) |E/E'| where ω = 0…..1. In this the large value of ω leads to more edge deletion. Cost of anonymization is minimized by integer programming and find optimal solution for small network.

4.1.1 Integer programming method

Integer programming finds the optimal solution for social network from friendship attack. It use following binary variable as: $\alpha_{u,v}$ denote the new edge added to G. $\delta_{u,v}$ is for an existing edge deleted from G. $\gamma_{u,m}$ is the degree of vertex u in m. Δ is set of degrees. $\varphi_{m,n}$ indicate if $D^2 = (m, n)$ is an anonymous group that needs to be protected from friendship attacks. Δ^+ is a set of positive degrees, $\theta_{m,n,u}$ is vertex u having degree m and is in an anonymous group $D^2 = (m, n)$. $\varepsilon_{u,v,n}$ is an edge (u, v) exists in the solution graph G, with the degree of vertex v as n. and for the integer programming the objective function are: $\min \omega \sum_{(u,v)\notin E} \alpha_{u,v} + (1-\omega) \sum_{(u,v)\in E} \delta_{u,v}$

Three constant is giving by integer programming for k^2-degree anonymity

4.1.1.1 Degree constraints

$\sum_{(u,v)\notin E} \alpha_{u,v} + (1-\omega) \sum_{(u,v)\in E} \delta_{u,v} = \sum_{m\in\Delta} m \times \gamma_{u,m}, \forall u \in V$

It finds the degree of each vertex where $\sum_{(u,v)\notin E} \alpha_{u,v} + (1-\omega) \sum_{(u,v)\in E} \delta_{u,v}$ is the sum of number of edges added and the number of existent edges that are not deleted, $\sum_{m\in\Delta} m \times \gamma_{u,m}, \forall u \in V$ says that $\gamma_{u,m}$ with the correct degree 1 for every vertex u.

$\sum_{m\in\Delta} \gamma_{u,m} = 1, \forall u \in V$ is ensure that only unique degree is selected for each vertex.

4.1.1.2 Anonymization constraints

To identified anonymous group that is to be protected is define by:

$\gamma_{u,m}+\gamma_{v,n}+\alpha_{u,v} \leq 2+\varphi_{m,n}, \forall (u, v) \notin E, \forall m, n \in \Delta^+$

$\gamma_{u,m}+\gamma_{v,n}-\delta_{u,v} \leq 1+\varphi_{m,n}, \forall (u, v) \in E, \forall m, n \in \Delta^+$

k^2-degree anonymization by at least k vertices for each of the two degrees in an anonymous group is given by

$k \times \varphi_{m,n} \leq \sum_{u\in V} \theta_{m,n,u}, \forall m, n \in \Delta^+$ and the degree of each selected vertex u must be m is giving by $\theta_{m,n,u} \leq \gamma_{u,m}, \forall m, n \in \Delta^+, \forall u \in V$.

4.1.1.3 Enforcement constraints

It gives each chosen vertex u with degree m connect to another vertex with degree n with $\theta_{m,n,u}$ is 1.

$\theta_{m,n,u} \leq \sum_{v\in V:v\neq u} \epsilon_{u,n,v}, \forall m, n \in \Delta^+, \forall u \in V$, it give at least one vertex v with $\epsilon_{u,n,v} = 1$ as result.

$\varepsilon_{u,v,n} \leq \gamma_{v,n}, \forall u, v \in V, \forall n \in \Delta^+$, it give the degree of v is n.

$\varepsilon_{u,v,n} \leq \alpha_{u,v}, \forall (u, v) \notin E, \forall n \in \Delta^+$,

$\varepsilon_{u,v,n} \leq 1 - \delta_{u,v}, \forall (u, v) \in E, \forall n \in \Delta^+$

Both above equation consider u and v must be connecting by an edge in G. If the edge does not appear in the input graph G, $\varepsilon_{u,v,n} \leq \alpha_{u,v}, \forall (u, v) \notin E, \forall n \in \Delta^+$ add this edge. $\varepsilon_{u,v,n} \leq 1 - \delta_{u,v}, \forall (u, v) \in E, \forall n \in \Delta^+$ enforces that this edge cannot be deleted, i.e., δu,v = 0, when (u, v) is in G.

4.1.2 Scalable approach

Scalable approach is use for k^2-degree anonymization of large-scale network. For this, degree sequence anonymization algorithm (DESEAN) is use. In this, firstly cluster vertex with similar degree where each cluster contain at least k vertex. After that add or delete edges and finally adjust the edges in the graph so that all vertex in each cluster satisfied the target degree.

4.1.2.1 Degree sequence anonymization

Decreasing order of the degrees, i.e., $d_{vi} \geq d_{vj}$for $\forall i \leq j$ and cluster them with similar degree group. After clustering vertices evaluates the cost CluCost (v_i, v_j, d^x) of grouping vertices $v_{i\ldots\ldots\ldots}v_j$in the same subsequence with a target degree d^x.

CluCost (v_i, v_j, d^x) = $(1 - \omega)\sum_{d_{vi \geq d_u} > d^x}(d_u - d^x) + \omega \sum_{d^x > d_u \geq d_{vj}}(d^x - d_u)$, this anonymization cost is calculate from the number of edges removed and deleted.

Where $\sum_{d_{vi \geq d_u} > d^x}(d_u - d^x) = \left(\sum_{d_u > d^x} d_u - \sum_{d_u > d_{v_i}} d_u\right) - d^x(LV[d^x] - LV[d_{v_i}])$ = Dec2Deg $[d^x]$ − Dec2Deg$[d_{vi}]$ − $(d_{vi} - d^x)$LV $[d_{vi}]$

Where LV[d] = number of vertices with degree larger than d. and Dec2Deg [d] =$\sum_{d_u > d} d_u - d \times$ LV [d] and

$\sum_{d^x > d_u \geq d_{vj}}(d^x - d_u) = d^x(SV\ [d^x] - SV\ [d_{vj}]) - (\sum_{d_u < d^x} d_u - \sum_{d_u < d_{v_j}} d_u)$ = Inc2Deg$[d^x]$ − Inc2Deg$[d_{vj}]$ − $(d^x - d_{vj})$SV $[d_{vj}]$.

Where, SV[d] = number of vertices with degree smaller than d. and Inc2Deg[d] = $d \times SV\ [d] - \sum_{d_u < d} d_u$

Now DegCost (v_1, v_j) is the minimum cost of dividing the overall vertex sequence v_1, ..., v_j so that each cluster contains at least k vertices. DegCost (v_1, v_j) derive from dynamic programming as:

$$\text{DegCost}\ (v_1, v_j) = \min \begin{cases} \min_{d^x} \text{CluCost}\ (v_1, v_j, d^x) \\ \min_{k \leq i \leq j-k}(\text{DegCost}\ (v_1, v_i) + \min_{d^x} \text{CluCost}\ (v_{i+1}, v_j, d^x)), \end{cases}$$

4.1.2.2 Privacy constraint satisfaction

In privacy constraint, satisfactory edges added or removed between the cluster pair(x, y). The number of edges connects with vertex in x and y may be zero or not less than k. For edges added or deletion, it maintain two table SatTab[x][y] for number of vertices in x with edges connecting to the vertices in y and DelTab[x][y] for number of vertices in x with edges disconnect to the vertices in y. Anonymization cost from edge deletion, is derived by $(1 - \omega)$ DelTbl[x][y],for pair of clusters (x, y), when 0< SatTbl[x][y] < k or 0 < SatTbl[y][x]< k and the anonymization cost from edge addition is derived by ω(k − min(SatTbl[x][y], SatTbl[y][x])).

4.1.2.3 Anonymous degree realization

In Anonymous degree realization, readjust the edges to achieve the target degree for each vertex $u \in x$ according to their degree order. If the degree of u is smaller than the target degree d^x, a subsequent vertex $v \in y$ is determine with the largest degree difference $d^y - d_v$, when SatTbl[x][y] and SatTbl[y][x] are not zero. After that an edge (u, v) is added to increase the degree of u without violating k^2-degree anonymity for protecting the pairs of connected vertices in x and y.

4.2 Protecting from structural attack

Identity and link privacy of user's throw structural attack is protected by k-automorphism [7]

4.2.1 K-automorphism

A graph G is k-automorphism if there exist k-1 automorphic functions F_a (a=1... k-1) in G, and for each vertex v in G, $F_{a1}(v) \neq F_{a2}(v)$ ($1 \leq a1 \neq a2 \leq k - 1$). There are no structural differences between v and each of its k-1 symmetric vertices, so attacker cannot find v from the other k-1 symmetric vertices having any structural information with a probability higher than 1/k. All the link information of users is preserve in anonymized graph and a systematic method is used to generate a k-automorphism graph. K-match of k-automorphism has some limitation as, if there are k matches in generating query then no one is identified the target in the network. However, k-automorphism is not sufficient if users have some shared vertexes and privacy of these shared vertexes still be disclosed. This limitation is removed by different matches, which are:

Different matches:- In a Subgraph query, if two arbitary matches m_1 and m_2 matches in the network and are isomorphic to function f_1 and f_2. If no vertex match with m1 and m_2 then both of m_1 and m_2 is different. In addition, k-different match principal use for protecting identity disclosure of anonymize graph generate after k-match algorithm.

K-different match- A release network after k-match and a sub-graph query, if there exist at least k matches of query in k-match and two of the k member is different matches.

4.2.2 K-match algorithm

This algorithm gives an anonymized network G^* for satisfaction of k-different match principal and take a guarantee of privacy through structural attack. In this algorithm, the identity of network is preserved by replacing attribute value with a random arbitrary value and get anonymized graph G'. Then this anonymized graph is partitioning into n block, these blocks are cluster into m groups U_i and each group has at least k blocks. Original blocks are replaced by alignment block, which is obtained by block alignment operation. After that edge copy operation is performed and get anonymization graph G^* generated after k-match algorithm. In block alignment operation each block P_{ij} , j= 1….k, in U_i is transformed into alignment block P'_{ij}, where all P'_{ij} are isomorphic to each other and must be k-1 symmetric vertices in other k-1 blocks. Group containing v and v's symmetric vertices from alignment vertex I = k.

4.2.3 Graph partitioning

Graph partition and block clustering lead to different anonymization cost and this should be kept minimum. Graph is partitioning into n block, these blocks are cluster into m groups U_i and each group has at least k blocks. Then anonymized cost is Cost (G, G^*) = $\sum_{j=1}^{k} \text{Cost}\ (U_i)$. Cost(G, G^*) = $(E(G) \cup E(G_*)) - (E(G) \cap E(G^*))$ and Cost (U_i) = AlCost (U_i)+0.5 * (k−1) * $\sum_{j=1}^{k} |\text{CrossEdge}(P_{ij}\ )|$. Graph partition firstly find sub graph R = {g_f} with the largest number of edges by setting min_sup = k. All sub-graph groups are cluster in one group and block are denoted by P_{ij}. If AlCost(Ui) = 0, then blocks are isomorphic to each other. If AlCost (Ui) $\neq$ 0 then dummy vertices need for alignment, and if AlCost (Ui) is small, then vertices are inserted into block have high structural similarity and if a vertex v contains more than one neighbour then it randomly inserted one of them. Few dummy vertex is inserted and alignment of then is based on their vertex degree. If only few edges are introduce between blocks then cost of anonymization graph is reduce and find |E (G') | = 0, |E (G') | ≥ k.

4.2.4 Block alignments

This algorithm gives aligned block and constructs AVT, which contain all alignment vertexes. According to AVT, firstly determine k vertex of a same degree from k block P_{ij}. If there is multiple choice of d then choose largest value of d and if there is no vertex in one block P_{ij} with the same degree as the selected vertices in other blocks, then choose the largest vertex in P_{ij}. After this alignment, breath-first search initiate in each block and pair up same degree vertices. If same degree vertices cannot found in other blocks, then dummy vertices introduces with same label of degree. Block alignment gives k alignment blocks P'_{ij} and another anonymized graph G".

4.2.5 Edge copy

The edges creation between two blocks in G" is a process to generate k-1 automorphic function F_a, which gives actual anonymized graph G^*. Function F_a is based on AVT and alignment vertex instance I treat as circularly linked list. This

F_a function give anonymized graph by copy edges in which if a vertex has one neighbour outside of block P then it is boundary vertex and if two boundary vertex of different group is connect with an edge then this edge called crossing edge.

4.2.6 Vertex ID generalization

This is used when anonymized network is repeat periodically and protect from intruder when he know the sequence of data released [15, 16]. For dynamic release network, find the connection between different release networks for ensuring privacy in this network. In vertex id, the entire vertex degree keeps same in different released of a same network. This algorithm is depend on two AVT A_1 and A_t and anonymized graph G_1^* and G_t^*, which obtain by k-match algorithm at time T_1 and T_t. after that two automorphic function F_a^1 and F_a^t are define based on A_1 and A_t in G_1^* and G_t^*. For every vertex v, if F_a^1(v) $\neq$ F_a^t (v), then vertex ID F_a^1(v) inserted into F_a^t (v)' vertex ID and vertex ID is replace by generalized ID for getting anonymized graph after vertex ID generalization.

5. COMPARATIVE EVALUATION

5.1 Privacy breaches under these attack

Both type of attack and their methods were compared and protect the released network from structure and friendship attack and ensure that no vertex in network is distinguish from other k-1 vertex with a probability higher than 1/k. The methods that have proposed for privacy preserving in social network are based on identity and link disclosure. Both the methods were compare in terms of cluster coefficient and average shortest path length. In these privacy methods, value of clustering coefficient and average shortest path length is taking for every value of k = 5, 10, 15, 20 and 25.

5.2 Data utility evaluation

Both method were compared on the basis of cluster coefficient and average path length. This comparison of method shows that graph anonymization cost is reduced by minimizing the addition and deletion operation on the network.

5.2.1 Clustering coefficient

It is use to measure the degree of node in graph which tend to cluster together. In social networking, node creates close groups. It is the ratio of all pairs of adjacent edges, which are completed by a third edge to form a triangle. The clustering coefficient varies between 0 and 1. Higher value of network shows the fully clustered network.

5.2.2 Average path length

Shortest path is one of the compulsive measures. Shortest path length sunders an easily conveyable network from one, which is strait and inefficient. Shorter path length is more desirable. This shorter path length is use for fast transformation and reduces the cost of the network.

K^2- degree anonymity work with vertex degree and friendship attack on the data set and evaluate the percentage of vertices that could be revealed with a probability greater than 1/k. According to this, the friendship attack is more harmful than vertex degree attacks. Power of friendship attack is trial on data set by scalable approach (DESEAN) with value of ω = 0.5.

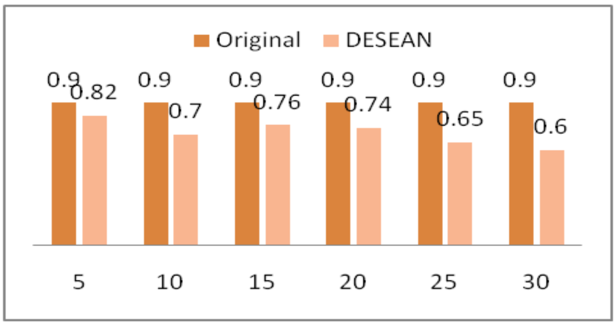

Figure 3: clustering coefficient of DESEAN

In figure 3: the value of original network is constant but the value DESEAN is continually decrease after anonymize the graph. In figure, the value of cluster coefficient decrees 0.82 to 0.6, when value of k is small then cluster coefficient is 0.82 and when k is large then it is 0.6. Increasing value of k gives the decreasing value of cluster coefficient.

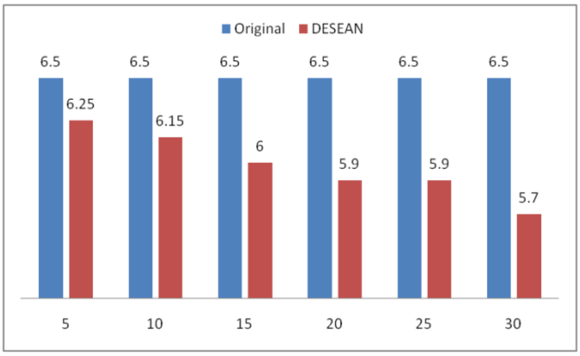

Figure 4: Average path length of DESEAN

Figure 4 gives a average path length of a given data set. Average path length of network continuously decrees to 6.25 to 5.7 and this decreasing value is desirable for minimizing the network cost. This average path value says that graph has performed minimum number of edge addition and deletion operation.

Whereas k-automorphism method also apply on real data set and generate cluster coefficient and average shortest path length graph for protecting the privacy throw structural attack and minimized the cost of anonymized network. It protect from vertex re-identification and structural attack in which intruder has some prior knowledge about network structure.

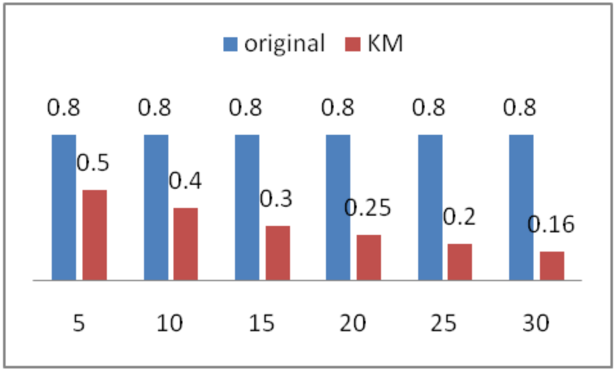

Figure 5: Cluster coefficient of k-match

In figure 5: the value of original network is constant but the value of k-match is continually decreases after anonymize the graph. In figure, the value of cluster coefficient decrees from 0.5 to 0.16, for smaller value of k the value of cluster coefficient is 0.5 and for larger value of k the value set to 0.16. Increasing value of k gives the decreasing value of cluster coefficient.

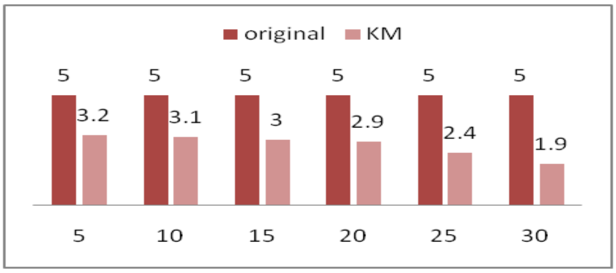

Figure 6: Average path length of k-match

Figure 6 gives a average path length of a given data set. Average path length of network continuously decrees to 3.2 to 1.9 and this decreasing value is desirable for minimizing the network cost and has minimum number of addition and deletion operation.

5.3 Result

The given two methods is preserving the privacy of user as identity and link information of users. This method gives different values of cluster coefficient and average path length as the value of k increase. Both of cluster coefficient and average path length give desirable result of privacy when values are small. It means the smallest value gives the most valuable solution for protecting the privacy of users throw intruder. Small value of average path length show that a minimum number of edge addition operation is perform on the anonymize network.

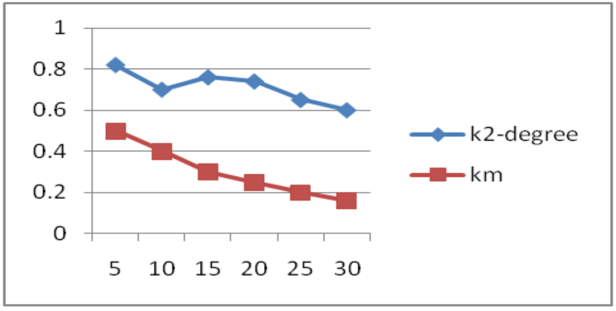

Figure 7: Comparison between clustering coefficient of DESEAN and K-match.

It show that k-match give more effective clustering coefficient value than the DESEAN anonymization. The cluster coefficient value generate by k-match are 0.52, 0.4, 0.32, 0.25, 0.2 and 0.16 and generate by k^2- degree are 0.82, 0.7, 0.76, 0.73, 0.65 and 0.6 respectively when k = 5, 10, 15, 20, 25 and 30. In figure, it shows that clustering coefficient value give by k-match is minimum compare to k^2- degree.

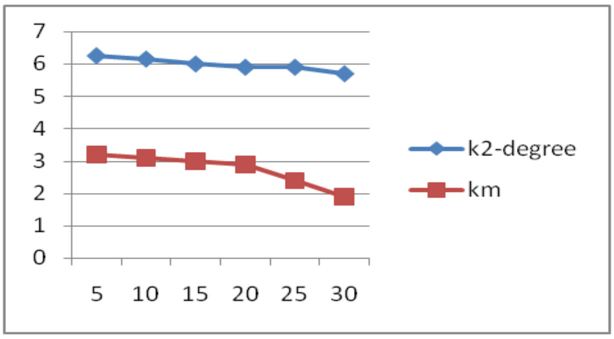

Figure 8: Compression between average path length of DESEAN and K-match.

It show that K-match give more effective average path value than the DESEAN anonymization. The average path value generate by k-match are 3.2, 3.1, 3.0, 2.8, 2.4 and 1.9 and value generate by k^2- degree are 6.25, 6.15, 6, 5.9, 5.8 and 5.6 respectively when k = 5, 10, 15, 20, 25 and 30. In figure, it shows that average path value given by k-match is minimum compare to k^2- degree.

6. CONCLUSION

Privacy preserving is one of the most important issues in the field of social networking. In this paper two algorithm namely k-automorphic and k^2-degree algorithm proposed by Lei Zou and Chih-Hua Tai are compared. Both the algorithm were tested on a real dataset considering the cluster coefficient and average path length as the performance metrix and it has been found that the k-automorphic show much better result in comparision to k^2-degree algorithm this is because K-automorphism gives small cluster and average path length value. In future, we extend and compare the performance of algorithm with parameter like degree distribution, degree centrality and number of edge change for preserving the privacy in the social network sites.

REFERENCES

[1] C. C. Aggarwal (Ed.), social network data analysis, DOI 10.1007/978-1-4419-8462-3_10, @ Springer Science+Business Media, LLC 2011.

[2] Fard, Amin Milani, Ke Wang, and Philip S. Yu. "Limiting link disclosure in social network analysis through Subgraph-wise perturbation." In Proceedings of the 15th International Conference on Extending Database Technology, pp. 109-119. ACM, 2012.

[3] Tai, C. H., Yu, P. S., Yang, D. N., & Chen, M. S. (2011, August). Privacy-preserving social network publication against friendship attacks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1262-1270). ACM.

[4] Tai, Chih-Hua, S. Yu Philip, De-Nian Yang, and Ming-Syan Chen. "Structural Diversity for Privacy in Publishing Social Networks." In *SDM*, pp. 35-46. 2011.

[5] Cheng, James, Ada Wai-chee Fu, and Jia Liu. "K-isomorphism: privacy preserving network publication against structural attacks." In Proceedings of the 2010 international conference on Management of data, pp. 459-470. ACM, 2010

[6] Zhang, Lijie, and Weining Zhang. "Edge anonymity in social network graphs." InComputational Science and Engineering, 2009. CSE'09. International Conference on, vol. 4, pp. 1-8. IEEE, 2009

[7] Zou, Lei, Lei Chen, and M. Tamer Özsu. "K-automorphism: A general framework for privacy preserving network publication." Proceedings of the VLDB Endowment 2.1 (2009): 946-957.

[8] Hay, Michael, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis. "Resisting structural re-identification in anonymized social networks."Proceedings of the VLDB Endowment 1, no. 1 (2008): 102-114.

[9] Korolova, Aleksandra, Rajeev Motwani, Shubha U. Nabar, and Ying Xu. "Link privacy in social networks." In *Proceedings of the 17th ACM conference on Information and knowledge management*, pp. 289-298. ACM, 2008.

[10] Liu, Kun, and Evimaria Terzi. "Towards identity anonymization on graphs." In*Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pp. 93-106. ACM, 2008.

[11] Zheleva, Elena, and Lise Getoor. "Preserving the privacy of sensitive relationships in graph data." In Privacy, security, and trust in KDD, pp. 153-171. Springer Berlin Heidelberg, 2008

[12] Zhou, Bin, and Jian Pei. "Preserving privacy in social networks against neighborhood attacks." In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pp. 506-515. IEEE, 2008.

[13] Machanavajjhala, Ashwin, et al. "l-diversity: Privacy beyond k-anonymity." ACM Transactions on Knowledge Discovery from Data (TKDD) 1.1 (2007): 3.

[14] N., Li, T., & Venkatasubramanian, S. (2007, April). t-closeness: Privacy beyond k-anonymity and l-diversity. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 106-115). IEEE

[15] Backstrom, Lars, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. "Group formation in large social networks: membership, growth, and evolution." In*Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 44-54. ACM, 2006.

[16] Kumar, Ravi, Jasmine Novak, and Andrew Tomkins. "Structure and Evolution of Online Social Networks." (2006).

[17] Samarati, Pierangela. "Protecting respondents identities in microdata release."Knowledge and Data Engineering, IEEE Transactions on 13, no. 6 (2001): 1010-1027.