International Journal of Security and Its Applications
Vol.9, No.8 (2015), pp.195-204
http://dx.doi.org/10.14257/ijsia.2015.9.8.16

Sensitive-resisting Relation Social Network Privacy Protection Model

Han Yan

Engineering Training Center, Inner Mongolia University of Science &Technology, Baotou, China
hanyannkd@sina.cn

Abstract

The existing social network privacy protection method mostly aims at the individuals of the social network, which cannot protect effectively the sensitive relations in the social network. Therefore, this paper proposes a new personalized K_L model. This model requires each sensitive relation with the sensitive relational point have l at least, and also the point with the same requirement has k at least. Thus, the attack has been resisted during the protection of the sensitive relations. Through seeking the most figure of merit sequence and considering individual sensitive attribute, the L-diversity method is applied so as to guarantee the least side and reduce the anonymous cost. Through the data set experiment, this paper proposes new personalized model K_L, which has the high anonymous quality and can effectively protect user's privacy in the social network.

Key words: *Social network; Privacy protection; K_L; Personalization*

1. Introduction

Nowadays, Social Network plays a more and more important role in the fields such as the analysis of Social psychology, disease research, market forecast and so on. However, as a result of social network containing a large number of individual information, its distribution and analysis will threat individual privacy, so simply deleting individual's identification information in a social network cannot effectively protect the privacy of individuals. Because an attacker can re-identify the node corresponding to the individual based on the other properties of the information or network structure characteristics in social network node, many institutions started research related to the technology about social network security release.

At present, the research work mainly focus on the perspective of attackers after the release of the social network and analysis of the protecting privacy in social network with different methods of anonymity. From the perspective of attacker, the attacker forms, in general, can be divided into property attack and structure attack. Property attack is that individuals of social network are identified based on the attribute information in social network node. Structure attack is that individuals are identified based on structure characteristics of social network. Structure attack mainly includes degree attack, neighborhood attack and subgraph attack, among which degree attack is one of the most common attack that an attacker uses frequently. Hay et.al[1] further found that the structure similarity degree of the neighbor node decides the degree of the individual who is identified in the network. Structure information and the degree of nodes are closely related to their neighbors. According to this, the author proposes a social network model which meets k-candidate anonymity. If you query each structure on the map, there exist at least k nodes that match the query. Structure query checks neighbors which the node exists or sub-graph structure which the node is adjacent to. Most of these authors focus on providing anonymous definition sets and studying their properties, but they do not pay

ISSN: 1738-9976 IJSIA

attention to the design of their anonymity algorithm that meet the requirements of the graph structure. Some properties of nodes (such as node degree) for individuals belong to sensitive privacy. In view of degree attack, the literature review [2] had proven that a social network chart apex can be distinguished by background knowledge of the chart structure even if there is no any label. For example, it's goal apex. For resisting degree attack, the literature review [2] proposed a standard of k-degree anonymous chart, which is similar to k-anonymous in the relation data. Specifically, the graph is a K-degree anonymous graph if V in each node of the graph has the same degree with at least (k-1) nodes. The probability of an attacker using degree background knowledge to re-recognize the target identity is up to 1/k. In order to solve the problem of degree anonymous in literature review [2], in literature review [3] the problem of subset of anonymous in social network was considered based on the subgraph of degree-constrained. In recent years, concerns about privacy are becoming increasingly prominent in the social network. Anonymizing a meaningful figure is a challenging problem because the original property of the figure must also be well protected. For this problem, in the literature review [4] its goal is to anonymizing subsets node, and, at the same time, ensure the minimum number of edges to add as far as possible. The main contribution of the literature review [4] is to design an efficient algorithm for the study of this problem by exploring it and subgraph problem of constraint to degree. Chester [5]et al. use an arbitrary inputting graph to give an algorithm producing k degree anonymous chart based on the dynamic programming. Recently, Yuan[6] et al. have introduced the subset anonymous ideas. In other words, it is not necessary to anonymize entire network because different users have different degrees of care regarding the privacy. Similarly, some vertices may not need privacy, such as films and film comments. The concept of subset of anonymous with label map is regularized by Chester et al. The literature review [7] also shows in the actual implementation of the many variants in a subset of the anonymous is a non deterministic polynomial difficult. In view of attribute attack, A. Campan and other scholars proposed the k- anonymous model, and this model requires there exist more than k-attribute individual which cannot be differentiated in social network. In order to resist the neighborhood attack, B. Zhou[8] proposed the k- neighborhood anonymous model, and this model requires there exist at least k-1 individual for any individual in social network, which can make sub-graph that is constituted by the neighborhood of these individuals be an isomorphism. L. Zou[9] proposed k- automorphism anonymous model. In the model, for any individual of social network , there exist at least k-1 individuals with the same information structure of others, which can resist various types of structure attack.

The above models mainly take account of is the individuals in social network, but sensitive relationships in social network are not considered. Sensitive relationships in social network are ubiquitous. If these relationships are not protected, individual privacy information will be let out such as the relationship between male and female friends, the individual and the organizations, the individual and the family. These individuals may be unwilling to divulge individual privacy information. Lan Lihui[10] proposed the privacy protection of the sensitive side, and designed the (k, 2) - anonymity algorithm. But his\her anonymity algorithm does not solve the problem of structure attack.

According to the degree that aggressor has and the background knowledge of sensitive relationships, this article designs a K_L personalization privacy protection model. In this model, based on users' different demands, there exist at least k individuals with the same degree in the anonymous network, which can cause probability of using degree attack to recognize the target to be lower than 1/k; Or there exist at least k individuals with the same degree and at the same time there is at least sensitive relationship in sensitive relational node, which causes the probability that the aggressor mark definite has the sensitive relational point with this point to be less than 1/l. Besides, this article put forward anonymous methods to realized the K_L- anonymous model.

2. Personalization K_L Model

2.1 Basic Concepts

Definition 1: social network graph. A social network graph G = (V, E) is made up with the node set V and the edge set E. Node set v represents individuals in social networks and edge set e represents the relationship between individuals.

This article will study the social network graph containing sensitive relationships with the following characteristics: 1) individuals in a social network are of the same type; 2) the edge relationships between individuals in social network are various, and the edge is free of label and weights, among which a kind of relationship that need to be protected is called sensitive relationships.

Definition 2: social network with the sensitive relationships. The node with sensitive relationship social network G consists of a group with 4 tuples, and its form is G = (V, E, L, Lv, LE, T):

V stands for node collection, V = (v1, v2, v3,…, vn);

E stands for the collection, E = ((vi, vj, t)), in which i, j = 1,2,3…, n.

The node vi stands for the individual in social network; (vi, vj) stands for the relationship between vj

L stands for the label collection, which is attribute set between the node and the side.

Lv stands for the node label function, which is the node to its label mapping.

LE side stands for the label function, which is side to its label mapping.

T side stands for the type collection, T=(1,2,…, r, s), in which 1…r stand for the non-sensitive side type, and s stands for the sensitive side type.

A social network can easily be abstracted as a graph. Each vertex in the graph represents each individual in the network, and an edge represents the relationship between the two individual.

Definition 3: the degree of a vertex. The degree di of a vertex vi refers to the number of other vertices which are connected to the vi. $\{vj \subseteq V: (vj, vi) \in E, i \neq j\}$.

Definition 4: the degree sequences of a vertices. A degree sequence DSm = <d1, d2, …, d|m|>, is made up with each v in m degrees, among which is one set of $m \subseteq V$. The hypothesis is that the vertices that were assigned were arranged according to the descending order of the degree so that degree sequence that is formed is listed and classified in descending order.

Define a 5 k anonymity sequence. For a degree sequence of DSm[i, j] if k is anonymous, if K is anonymous, the number of each di's occurrence is at least k times in the DSm[i, j], of which $i < j$, i, $j \in m$, $m \subseteq V$.

Define the 6 k degree sequence of a node set. If frequency of occurrence of each the different values in the degree sequence of a graph are at least k times, this degree sequence is called the k-degrees sequence.

As Machanavajjhala[11] shows, in a relationship table, if the user's sensitive attributes lack diversity, k-anonymity does not adequately protect users ' privacy. Similarly, k-anonymity's social network data will still lead to divulging of privacy. If a node in an equivalent group or a sensitive property value in the property list of the edge are equal, the attacker will identify the sensitive attribute.

Based on principles of relationship tables l-diversity[11], the attacker with background knowledge of a vertex's degree can infer target users with sensitive relationship with the probabilities of less than 1/l. With the increasing of 1, privacy protection rises.

If an attacker can identify the relationship between two individuals based on the sensitive relationship in social network, the privacy of two individuals in network is released. For example, figure 1 stands for friend social network, in which nodes represent individuals and edges represent relationships. There are three types of edges: the dotted line represents the relationship between the students; the solid line shows the relationship

between colleagues; thick solid line stands for the relationship between male and female friends, assuming that their relationship is sensitive. In Figure 1, if an attacker, based on the background knowledge, knows node A is Jone and node E is Lily, he can accurately infer the friendship between Jone and Lily are boy and girl friendship from Figure 1, which may not be let out by Jone and Lily. What's more, the degrees of node-A-E in Figure 1 are 5, 4, 3, 3, 3. If the attacker knows that Jone's degree is 5, it is easy to identify node E was Mary by using the individual's number of degrees, which will let out her privacy.

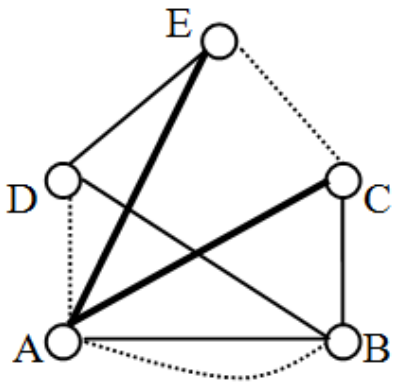

Figure 1. The Original Social Network

K_L- anonymous model is proposed here by constructing the k-degree sequence meet k- degree of anonymity for the protection of sensitive relationships.

Definition 2 (K_L- anonymous model) A graph G = (V, E, L, Lv, LE, T) is given: s is sensitive relationship in the diagram, s in T . For a node V in the V diagram, there exists at least k-1 nodes with the same degree as V; if the node V is related to s, there is at least l node which is related to V, then the node meets (k, l) - anonymity constraints. If every node in the graph can meet (k, l) - anonymity constraints, the graph meets (k, l) - anonymity model.

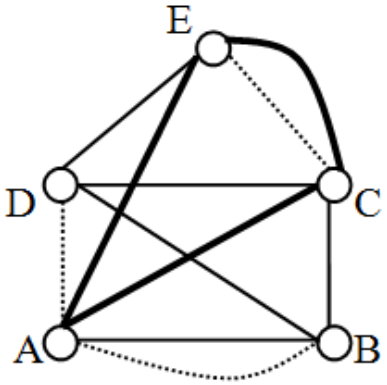

Figure 2. K_L Social Network Anonymous Graph

For example, Figure 2 is an anonymous network of figure 1 (k, l) in which the values of both k and 1 are 2. From figure 2 it can be seen that there exist at least 2 relationships if there is a sensitive relationship in any node. If an attacker knows the node C is Tom and node E is Lily, the probability that he or she can accurately infer the relationship between Tom and Lily decreases by 50 percent from figure 1. When the value of "l" increases, the probability that this sensitive relationship is accurately identified decreases. Besides, it can also be seen that there are at least 2 nodes with the same degree. If an attacker knows Jone's degree is 5, the probability that Jone is accurately identified decreases by 50 percent. As the value of "k" increases, the probability that the node is identified decreases.

This paper proposes K-L personalized model based on the above definitions.

3. Implementation of Personalized K_L Model

3.1 l-diversity Anonymity

Anonymity algorithm of adding sensitive sides is used to construct anonymous graph which contains sensitive sides. Sensitive sides are added circularly until the numbers of sensitive nodes which contain sensitive side node are at least l.

Theorem 1 (complexity of social networks l-diversity)[11] the problem of l-diversity in social networks is NP-hard.

For example: a social network with sensitive relationship G = (V,E,L,Lv,LE,T), l and n are integers.

Question: is there an anonymized social network G' = (V,E',L,Lv,LE,T) which makes such that E⊂E', | E' - E|≤ n, and G' is a l-divers?

Proof: We summed up the problem of k-anonymity in social networks as k-diversity in social networks. For a given social network G, if you want to meet its k-anonymization by adding at most n edges, we can create a new social network G' by assigning a unique sensitive label to each node in G. Obviously, if and only if there is a k-diverse anonymization G2 of G' and | E(G2) – E(G')|≤ n, there is a k-anonymization of G1 to get E(G1) and – E(G)|≤ n.

3.2 K -degree Anonymity

We minimize the anonymization cost by finding the k- degree anonymous graph. That is, adding a minimum number of edges to the original social network graph to form k degree anonymous graph.

Assuming L (d' - d) = $\sum_{i} | d'(i) - d(i) |$ represents the difference of degree sequence, the operation of minimizing edge is changed into minimizing the distance of L.

The social network graph is transformed into the form output of degree sequence of d so as to construct a k anonymity degree sequence of d', so we can get the degree anonymization cost:

DA (d ', d) = L (d' -d) is minimized.

Dynamic programming algorithm (DP)

The aim of dynamic programming algorithm is to let DA(d(G') ［1, n］) reach minimum, the state transition equation is:

DA(d(G') ［1, i］= I(d(G') ［1, i］) i ＜ 2k

DA(d(G') ［1, i］= $\min_{k\le t\le i-k}$ { DA(d(G') ［1, t］) +I(d(G') ［t + 1, i{］) } i≥2k (1)

finding the best k degree sequence.

Deep traversal of the existing graph and count the degree of vertex set. According to the calculated value of degrees, we rank them and they can be cited in the array of objects.

According to the array of objects that we get in the first step, we calculate the best k - degree sequence.

Using the ideas of dynamic programming in the k - anonymous algorithm, we look for the minimum cost of anonymous degrees sequence to form the best k - degree sequence.

Make a comparison of the new degree sequence and the original degree sequence, we give priority to add sides in the vertex set. Then, we consider to add sides to the vertex which is not in the vertex set. Therefore, the best degree sequence is formed.

The idea of dynamic programming is as follows:

a) Encode the node in the existing graph, extract subsets that need to be protected, get the statistics of the values of the degree, and sort a sequence formed from large to small.

b) For the sorted sequence of values, we get statistics of the number of times which each value appears, and compare each with k.

c) If it meets k-degree anonymization, we save all the relevant information. If it doesn't meet the condition of k, we calculate the cost of degree to find a suitable k. We define a container for t, and t meets some range to search the t value which make the degree consumption minimum in the range. First, record each t value. Second, record each t value and its cost. Third, if it has second tier of traverse, record a second-tier and even third-tier. With such a recursive call method, I use dynamic programming to find the best degree sequences.

d) If the current t value is less than or equal to k, we carried out the first part and second part consumption calculation. If it is larger than k, once again, calculate best degree consumption, that is, recursion into a layer.

e) Build a k anonymous graph by adding edges.

4. Anonymous Posting Algorithm

Based on the above discussion, the personalized K_L model and relevant algorithm are designed to prevent the recurrence of vertex recognition, and to prevent attacking the sensitive attribute consistency of vertex that led to privacy leak in this paper. The algorithms and analysis are as follows:

Input: social network G which contains sensitive relationship, An integer l and k
Output: k-degree sequence d of l- sensitive edge anonymous figure of (G *)

1. for i = 1 to | E | do
2. if (e [i] is a sensitive side) then
3. Sen [e [i] [1]] = Sen [e [i] [1]] + 1;
4. Sen [e [i] [2]] = Sen [e [i] [2]] + 1;
5. end if
6. end for
7. for i = 1 to n do
8. if (Sen [vi] ≠0) and (Sen [vi] < k) then
9. while (Sen [vi] < k) do
10. Random selection vj∈V, 且 vi≠vj, Sen [vj] ≠0,
11. if(vi, vj, s) ∉E then
12. E = (vi, vj, s) ∪E, Sen [vi] =Sen [vi] + 1, Sen [vj] =Sen [vj] +1,
13. end if
14. end while
15. end if
16. end for
17. output l- sensitive side anonymous graph G'.
18. for i = 1 to n do
19. for j = k to n do
20. $I(d(G') [i, j]) = \sum_{l=i}^{j} d(G') (i) - d(G') (l)$
21. end for
22. end for
23. for i = 1 to 2k-1 do
24. DA(d(G') [1, i]) = I(d(G') [1, i])

25. end for
26. for i = 2k to n do
27. DA(d(G') [1, i]) = $\min_{k \le t \le i-k}$ { d(G') [1, t] + I(d(G) ' [t +1, i] }
28. for j = t to i do
29. d(G*) (j) = d(G') (t) ;
30. end for
31. end for
32. output l- k- anonymous graph sensitive edge degree sequence d(G*)

5. The Experimental Results and Analysis

5.1 Experimental Environment

The hardware environment of experiment is: CPU Inter(R) Core(TM) i5(3.2 GHz), RAM 8G, and operation system is 64 bit Windows7, and the experiment tools are Eclipse 4.3.2, JDK 6.0. Test dataset is generated by Pajek software to build social network graph, which contains 200 nodes, with the average degree of 5.

5.2 Experimental Instructions

K value and l value are the main variables involved in the K_L model algorithm in this experiment. Experimental design 2 experiments.

In the first set of experiments, k is taken as 2, 4, 6, 8, 10, l=2, we contrast this algorithm(k_l) and classical k- algorithm (k_d) of anonymous information defect rate. The information damage rate can be measured through $r = \left(d - d^{*}\right) / d$. As shown in Figure 3, obviously, along with the k value increase, both algorithm damage rates all increase, but this article's algorithm surpasses the classics k- anonymous algorithm, and the algorithm performance is enhanced.

In the second set of experiments as shown in Figure 4, respectively, l take 2,3,4,5,6, k=5. The best 5 degree sequence is obtained in the original graph when k is 5. Respectively, based on the different L values, we compare the efficiency between this algorithm and the classical k- algorithm. As shown in Figure 4, it is clear that information damage rate in paper is low; data usefulness are improved compared with classic k-degree anonymous algorithm.

In the third set of experiments as shown in Figure 5, to test the utility of anonymized data, the original graph data and the anonymous data are used to measure properties of some graphs, such as the clustering coefficient (CC). As shown in the result, in the anonymous data, with the increase of K ratio, clustering coefficient decreased slightly. However, the clustering coefficient of the anonymous figure is still fairly close to the original data. Even when k = 10, the difference between clustering coefficients of the original graph data and the anonymous graph data is only 0.06.

In the fourth set of experiments as shown in Figure 6, the runtime on the same data sets are tested with respect to different k and l values. The runtime increases when the k and 1 value increase. Furthermore, the runtime for k_l algorithm is longer than that for k_d algorithm. There are two reasons. First, achieving l-diversity needs additional operations for l-diverse partition. Second, the cost for achieving l-diversity and k-degree is larger than that for achieving k-anonymity.

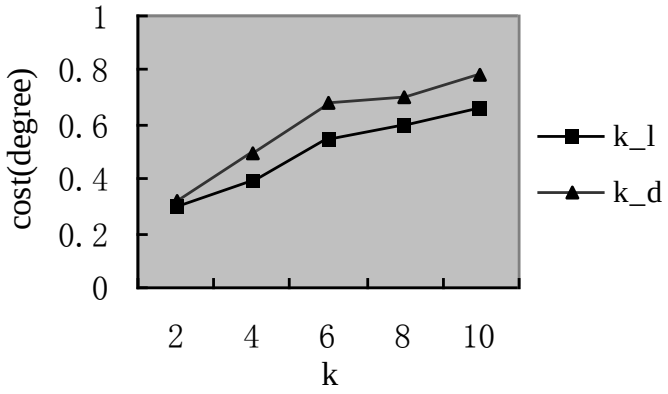

Figure 3. Degree Anonymization Cost with Different k

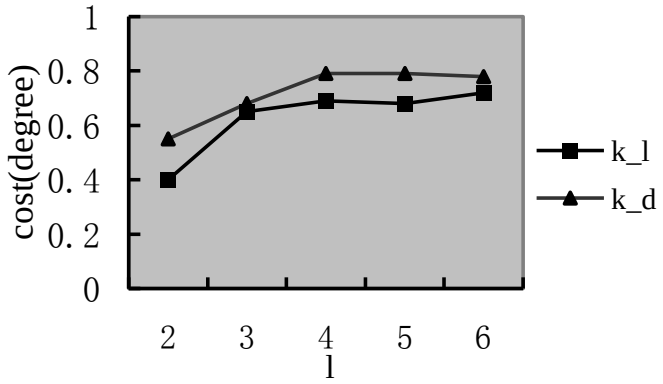

Figure 4. Degree Anonymization Cost with Different l

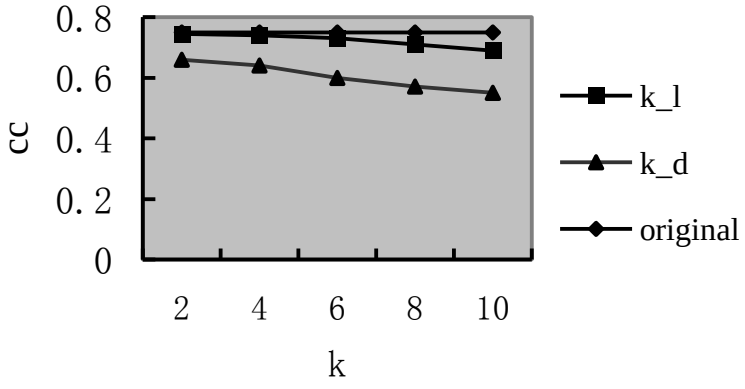

Figure 5. Degree Anonymization Cost with Different k

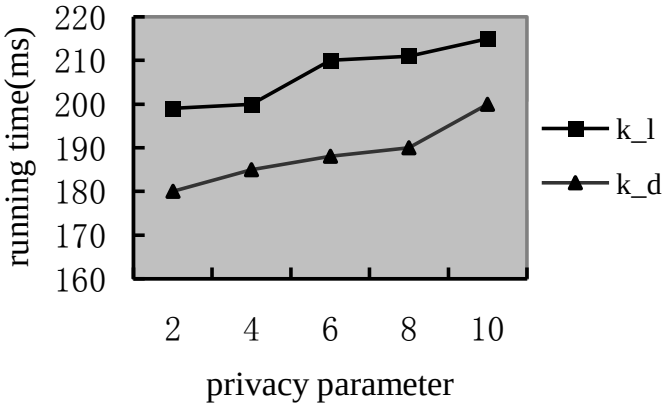

Figure 6. The Running Time of Different Algorithm

6. Conclusion

This paper proposes a social network personalized model to protect the data privacy. In the model, the strength of protection goes from weakness to strong, and recursive k _l anonymous algorithm and k _l anonymous algorithm are realized respectively. In order to realize k _l anonymous algorithm and minimize the cost of anonymity, this algorithm requires that the maximum degree of anonymous graph is less than or equal to the maximum degree of original graph, and construct new by adding the least sides between the nodes which are in original graph, and not adding new nodes. The experimental results proved that the model of the algorithm has higher efficiency and lower cost than the existing classic algorithms.

References

[1] M. Hay, G. Miklau and D. Jensen, "Resisting structural re-identification in anonymized social networks [J]", VLDB Endowment, vol. 1, no. 1, (**2008**), pp. 102-114.
[2] K. Liu and E. Terzi, "Towards identity anonymization on graphs [C]", In: Proceedings of the 2008 ACMSIGMOD International Conference on Management of Data (SIGMOD'08), New York: ACM Press, (**2008**), pp. 93-106.
[3] S. Chester, J. Gaertner, U. Stege and S. Venkatesh, "Anonymizing Subsets of Social Networks with Degree Constrained Subgraphs [C]", In: Proceedings of the 2012 ACM International Conference on Advances in Social Networks Analysis and Mining, New York: ACM Press, (**2012**), pp. 418-422.
[4] S. Chester, B. M. Kapron, G. Ramesh, G. Srivastava, A. Thomo and S. Venkatesh, "k-anonymization of social networks by vertex addition", In ADBIS 2012 Research Communications. Austrian Computer Society, (**2011**) September, pp. 107–116.
[5] M. Yuan, L. Chen and P. S. Yu, "Personalized privacy protection in social networks [J]", Proceedings of the VLDB Endowment, vol. 4, no. 2, (**2010**), pp. 141-150.
[6] S. Chester, B. M. Kapron, G. Srivastava, and S. Venkatesh, "Complexity of social network anonymization," Social Network Analysis and Mining, (**2012**) March.
[7] A. Campan and T. M. Truta, "A clustering approach for data and structural anonymity in social networks [C]", In Proceedings of the 2nd ACM SIGKDD International Workshop on Privacy, Security, and Trust in KDD, in Conjunction with KDD' 08, LasVegas, Nevada, USA, (**2008**), pp. 93-104.
[8] B. Zhou and J. Pei, "Preserving privacy in social networks against neighborhood attacks [C]", In Proceedings of the 24th IEEE International Conference on Data Engineering, Cancun, Mexico, (**2008**), pp. 506-515.
[9] L. Zou, L. Chen and T. Ozsu, "K-automorphism: a general framework for privacy preserving network publication [C]", In Proceedings of VLDB Conference, (**2009**), pp. 946-957.
[10] L.-h. Lan, Y.-h. Sun, S.-g. Ju, "Privacy pre-servation of sensitive edges in social networks publication [J]", Proceedings of Journal of Jilin University(Information Science Edition), vol. 29, no. 4, (**2011**), pp. 324-331.

[11] A. Machanavajjhala, J. Gehrke, D. Kifer and M. Venkitasubramaniam, "L-diversity: privacy beyond k-anonymity [C]", In: Proceedings of the 22nd IEEE international conference on data engineering (ICDE'06), IEEE Computer Society, Washington, DC, (**2006**), pp. 934-987.

Author

Han Yan, he received the BS in computer science from Inner Mongolia Agricultural University, China, in 2002, and received the MS in computer science from Inner Mongolia University of Science & Technology, China, in 2010. Currently, her research interests include Computer network technology, and concerns of the development of next generation networks and has accumulated a wealth of experience in network security, and also modern cryptography.

