Privacy Preserving in Social Networks Using Combining Cuckoo Optimization Algorithm and Graph clustering for Anonymization

Abstract

Recently, social networks have received dramatic interest. The speed of the development and expansion of the Internet has created a new topic of research called social networks or online virtual communities on the Internet. Today, social networking sites such as Facebook, Twitter, Instagram and so forth are dramatically used by many people. Since people publish a lot of information about themselves on these networks, this information may be attacked by the intruders, so the need of preserving privacy is necessary on these networks. One of the approaches for preserving privacy is the K-anonymity. Anonymization always faces the challenge of data lost, therefore, an approach is required for anonymization of data and meanwhile maintaining the usefulness of the data. In this research, by combining the k-anonymity priority clustering method and Cuckoo optimization algorithm, an appropriate model is developed to maintain the privacy of the data and its usefulness. The average path length, average clustering coefficient and the transitivity criteria have been used to evaluate the proposed algorithm. The results of the experiments show that the proposed method in most cases has 1 unit superiority in terms of k-anonymity and 2 units superiority in terms of usefulness in comparison with similar methods.

Keywords: Social networks, graph clustering, Cuckoo Optimization Algorithm, k-anonymity, utility of data,

1. Introduction

The speed of Internet development has led to a new research, called social networks on the Internet or virtual/online communities [1]. Nowadays, social networking sites such as Facebook, Twitter, Instagram, etc. are dramatically used by many people. Given that there is a lot of information about the users' privacy in social networking data, it is necessary to modify the actual data to ensure the privacy of users. If the correction is excessive, the utility of social networks' data is reduced [2]. On the other hand, if the correction is inadequate, the privacy data is not well protected, so, a balance must be created between privacy and utility [3].

In order to solve the problem of privacy violations of users, the systematic approach must be applied before the data are published. One of the methods of privacy in social networks is k-anonymity. One of the challenges in the anonymization is that the main graph features should be protected as much as possible, so that the utility can be maintained [4]. Another issue is the intranet background knowledge. An Intranet background knowledge creates substructure using the openness of social networks, which is easily known as basic knowledge before the publication of anonymous data. This substructure is usually a

particular vertex or subgraph. After the publication of anonymous data, the intranet can acquire sensitive information by identifying this structure [2].

An appropriate anonymization approaches are the k-anonymity approach that is suitable for dealing with the problem of background knowledge of the attacker [5]. In this research, at first, in order to reach the appropriate anonymity level, clustering of nodes has been addressed based on k-anonymity and then for anonymity of some clusters of graph which are not k-anonymity, we attempt to reach the number of edges and vertices to the minimum required for anonymity by Cuckoo Optimization Algorithm.

In the rest of the research, we will examine some of the performed work in Section 2. In the Section 3 a suggested method for preserving the privacy on the social networks are provided. The results of the evaluation of the proposed method and the discussion about it are presented in Section 4 and, finally, the conclusions and suggestions are given in Section 5.

2. Background of Research

Various studies have been conducted to eliminate concerns about privacy preserving and different mechanisms have been proposed. One of the mechanisms of the privacy preserving in social networks is anonymization, which we will discuss some of techniques and algorithms provided in this area.

Editing models of the random graph can reduce redefinition attacks, but with random editions of edges or vertices, this strategy ignores that the privacy must include all users. However, privacy assurances are provided only for some users randomly [6]. Hay et al. presented the simplest method of random graph for data anonymization of the social networks to protect the graph against the re-identification of the vertex and attack of edge disclosure. This algorithm removes the N-edge of the graph and adds the unreal M-edge, so that it is concluded that m = n [7]. In [8], a random approach of spectrum preserving is proposed, that is referred to removing/adding spectrum and switch of the spectrum. To maintain profitability, the special values are retained from both the adjacency matrix and the Laplacian matrix. Researchers claim that achieving a spectrum strategy is such as anonymization applied in [7], but spectral preservation leads to preserve many of the features of the graph.

In [9], researchers proposed two random sparsification and random perturbation algorithms. In the random perturbation, probability p is chosen by the algorithm and then the independent Bernoulli law B_e = (1-p) is calculated for each edge. Edges with B_e = 1 are deleted from the graph and edges with B_e = 0 remain unchanged. In the random perturbation approach, first, the edge $e \in E$ is deleted from the graph G with probability p, and then the edge e is added with probability q in $\binom{v}{2} \backslash E$. The probability q is defined by equation (1).

$$q = \frac{|E| \cdot p}{\left(\binom{V}{2} - |E|\right)}, \qquad (1)$$

In [9], the entropy probability distribution, has been used to determine the quantity of anonymity level. In this research, it has been shown that the mechanism of the random perturbation cannot achieve a high degree of anonymity without reducing the features of the graph. The authors proposed in [3] a new approach called division anonymization for preserving privacy. The division anonymization algorithm is divided into two sections: the anonymization of the vertex information and the division of the vertex. The first K-degree anonymization was proposed and evaluated by Liu and Tersey [5]. In this approach, the degree-sequence of the main graph becomes anonymity by the dynamic-programming approach, and the criteria of average distance and clustering coefficient are used to measure the information lost. Hartung et al. proposed a modified dynamic algorithm for k-anonymization of the degree-sequence [10]. This algorithm is a developed example of Liu and Tersey's work.

Researchers [11] proposed a k-anonymization based on genetic algorithm. In this approach, at the first, a degree-sequence is created in the main graph, and then a new k-anonymity sequence is calculated using the genetic algorithm. In [12], we tried to generalize the problem of the degree of anonymity proposed by Liu and Tersey [5]. To solve this problem, an algorithm based on k- anonymity is proposed, which is executed for a non-tagged network almost at O (nk) Polynomial-time. The researchers [13] proposed an anonymization approach with editing the edges and vertex to convert the social network to a k-anonymity version. A k-degree anonymity with vertex and edge modification algorithm (KDVEM) is designed to convert a graph to its k-anonymity version. Researchers [14] presented the theory of the k-Neighborhood Anonymity approach (k≥1), which transforms the main graph G into an anonymous version $\bar{G}$ for collective search processing. The k-Neighborhood Anonymity approach is a greedy graph modification algorithm that adds vertices and edges to a graph until a graph is generated isomorphic with the minimum of k-vertex with its neighboring sub-graphs.

Thompson proposed the concept of one-step of anonymity to address the issue of the privacy violation [15]. In this way, it is assumed that the enemy has previous knowledge of the vertex with degree of the target and its neighbors within the radius i. Here, the technique of matching internal clusters is developed one step to anonymization of the graphs against attacks by adding and removing the edge. In the study [16], the concept of the automorphism has been proposed to reduce the privacy violation based on the subgraph. This model guarantees the presence of the same minimum K-structural subgraphs published in the graphs. To minimize the information lost, the cost of anonymity is defined by equation (2).

$$Cost(G, \bar{G}) = \big(E(G) \cup E(\bar{G}) - E(G) \cap E(\bar{G})\big), \qquad (2)$$

Where *E (G)* is a set of edges in the graph G, and the lower cost indicates a lesser change to the main graph. Researchers [17] suggest the concept of k-symmetry, which assumes that the enemy has the prior knowledge of any graph that contains the person in question. This approach is based on the automorphism [16] and its aim is the anonymity of the subgraph level. Researchers [19] developed the concept presented in [16] by generating an isomorphic k-anonymous graph. The isomorphic k-graph called as $G = \{g_1, g_2, \dots, g_k\}$ provided that the graph is formed by the disjoint k-subgraph, so that $g_i$and $g_j$are both isomorphic and$g_i \neq g_j$. The authors [18] proposed a cluster-based anonymity technique for a simple unlabeled

graph. In this study, to measure of the utility lost, a sequence of graph feathers such as the relationship, the length of the shortest path and maximum degree are used.

A sequential clustering algorithm is proposed to anonymization a graph with higher utility in [20]. Sangny et al., with one of the new approach, provided the production of equivalent classes with minimum k-vertices [21]. In this way, each equivalence class can be considered as a cluster, and at the same time, the center of a cluster can be explained as a general form of an equivalence class. In the Vertex add method, the authors investigate the anonymization in a complete graph [22]. The basis of this method is greedy and does not guarantee the optimization of k-anonymity operations.

3. Suggested method

The concern of individuals about their privacy in these networks is also increasing due to the significant development of social networks, so a new approach for anonymity of the social network is presented that leads to the privacy preserving and maintains its utility. In this research, the social network is considered as a simple social graph with no direction $G = (V, E)$. The purpose of the proposed method is to minimize the information lost and converting the social network graph to k-degree anonymity graph.

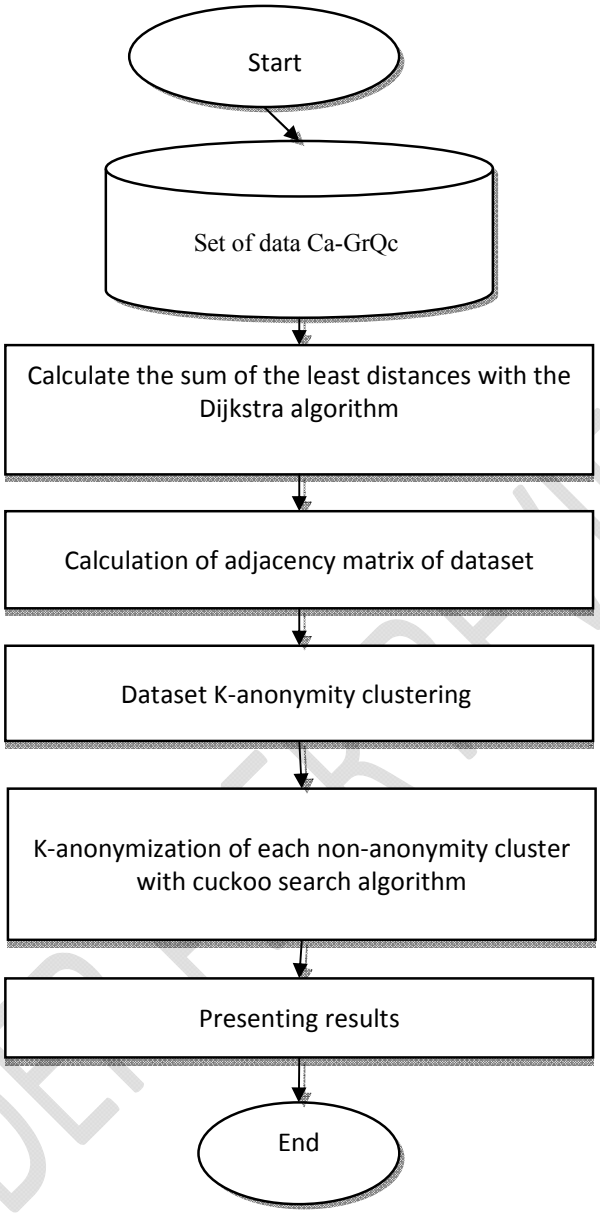

Figure 1: The proposed approach framework

To accomplish the determined goals, an appropriate solution for the k-anonymity of the graph is provided by combining the techniques of graph clustering and Cuckoo optimization algorithm that preserve the privacy. The proposed approach framework is shown in Figure 1.

In the proposed approach, at the first, the length of the shortest paths of vertices is obtained by the Dijkstra algorithm, and then the adjacency matrix is computed for the selected dataset. After calculating the adjacency matrix, the social network dataset reaches the anonymity. For anonymity, in the first step, the vertices are clustered to reach k-anonymity,

and then with the cuckoo optimization algorithm, k-anonymity of the graph data is completed. The steps of the anonymity of the proposed approach are described as follows.

3-1 Graph clustering

In this technique, in the first step, the vertices are clustered to reach k-anonymity. In clustering, the number of cluster members is important to reach k-anonymity. In each cluster, there are least k and maximal 2k-1 vertices with the same degree. In this research, clustering is applied based on the same degree and consideration of the closeness centrality. The clustering steps are as follows.

-Step One: The number of degrees of each vertex is obtained.

-Step two: Data is sorted according to degrees.

-Step Three: The closeness centrality of each vertex is calculated. The closeness centrality is calculated as the sum of the lowest paths from one vertex to the other available vertices. According to Equation (3), the closeness centrality of the matrix gravity is calculated from of the shortest path [10, 11].

$$C_c(v) = \frac{1}{\sum_{t \epsilon V} d(v,t)}, \quad (3)$$

Step Four: The vertices with the same number of degrees are incrementally ranked again according to the results of centrality.

Step Five: To reach the cluster at the k-anonymity level, k-vertex that has the same degree placed in a cluster. In this way, there are degrees more than the k-vertex with the same degree, the vertices that has, the less closeness centrality is placed in the cluster, and the vertices remain with a higher closeness centrality.

Step Six: At the end, if less than k-vertex remains in the graph, each one will placed in the nearest cluster in its own cluster. In this way, the difference of their degree is minimized and less variation takes place in the graph.

3.2 k-Anonymity of degrees with Cuckoo optimization algorithm

In the previous section, an approach was expressed for clustering the vertices to make the social network graphs as k-anonymity as possible. A number of clusters became k-anonymity despite having least k-degrees. Anonymous clusters are set aside and the remained clusters reach k-anonymity by adding and removing vertices and edges. During adding a vertex and an edge, the balance between the graph's anonymity and the utility of the data must be established. In other words, for anonymity, the number of vertices and edges will not be added too much that leads to eliminating the real data, and the amount of anonymization should not be so low that leads to the privacy violation by the enemy. For this purpose, Cuckoo's algorithm is used for optimization.

The main idea of using this algorithm is that by increasing or decreasing the degree and vertex in the cluster, all vertices in the cluster reach the same degree and the value of the

fitness function is also minimized. Steps for this algorithm to optimize and create the anonymity graph are as follows.

Step One: Initialization of the Cuckoo search algorithm parameters, which includes determining the primary population (n), the number of groups (m), determining the dimensions of the problem, the end condition (the number of the global searches), and so on.

Step Two: Cuckoo search algorithm with a primitive population that are randomly selected according to the dimensions of the created problem starts to do its task. The parameter of the number of vertices in the cluster and the degree of vertex parameter are considered as the cuckoo optimization algorithm's parameters.

Step Three: The fitness value for each cuckoo is equal to the difference between the sum of the degree cluster vertices in the main graph and the sum of the degrees of cluster vertices in the anonymity graph. The fitness function is calculated according to equation (4).

$$\Delta = \left| \sum_{i=1}^{n} d_C(v_i) - \sum_{i=1}^{n} d_{\tilde{C}}(v_i) \right|, \qquad (4)$$

The fitness of different states is added to the vertex up to twice the original number of vertices in each cluster. Adding and removing the edges is also considered for achieving the same degrees in each cluster.

Step Four: Cuckoos are arranged according to their fitness.

Step Five: Based on the steps of the algorithm, a new population is generated.

Step Six: The end condition is considered with respect to the number of main loop repeats of the algorithm. If the algorithm is repeated in the determined number, it goes to step 7, otherwise it goes back to step four.

Step 7: The optimal answer is obtained for the parameters of the number of vertices and the degree of vertices of each cluster.

3.3 Creating changes in the graph

By calculating the degree of each vertex in the main graph, the sequence of vertex and degree is obtained for each cluster. Based on the result obtained in the previous step, degree of each vertex is determined. Given the differences in the main graph sequence and the obtained size of the vertex and cluster, the changes are applied as follows.

- If it is required to add or remove vertex and edge, at the first, a vertex in a cluster after the last number of the vertex in the graph is added with zero edge and based on which vertex need the edge, one edge is connected to the extra vertex.

- Thus, the main graph is reconstructed and corrected.

4. Results and discussion

In this section, we examine the results of the implementation and evaluation of the proposed approach and comparing its performance with other similar approaches. MATLAB software version 2016a has been used to simulate and analyze the proposed approach. Simulation and all experiments were done using an Intel processor Core (TM) i7-6500u with the frequency of 2.5 GHz, 8 GB of memory and Windows 10 64-bit. In this research, the Ca-GrQc dataset has been used in experiments. GR-QC (General Relativity and Quantum Cosmology) is a collaborative network of archives Web for the pre-print edition, and covers scientific collaboration between the authors of papers presented to the Department of General Relativity and Quantum Cosmology. The Ca-GrQc data include papers from January 1993 to April 2003 (124 months), and represents the full history of the GR-QC section [23]. Table 1 shows the features of the Ca-GrQc dataset.

Table 1: Features of the Ca-GrQc dataset

Type	Number of Vertices	Number of Edges	Average the length of the path	The average of the clustering coefficient	Transitivity
No Direction	5242	14496	6.049	0.687	0.630

In order to evaluate the effectiveness of the proposed approach, three important criteria for analyzing social networks have been investigated [13]. The average path length (APL) is the distance between the two vertices (u, v) of the shortest path between u and v in the main graph and is calculated by equation (5).

$$APL = \frac{\sum_{(u,v)\in RP} SPL(u,v)}{|RP|}, \qquad (5)$$

In this regard, RP specifies all available pairs of vertices and SPL (u, v) the shortest path length between vertices u and v as well. This is used for information efficiency or transitivity volume in a network. The criterion of the average clustering coefficient (ACC) is the ratio of the possible triangle for the vertices and is calculated by the equation (6).

$$C_u = \frac{2T(u)}{\deg(u).(\deg(u)-1)}, \qquad (6)$$

In this equation, *T (u)* indicates the number of triangles through the node u and deg (U). The transitivity criterion is one of the clustering coefficients that specifies and measures local loops near a vertex. This transient criterion calculates the number of triangles and triangles according to (7).

4.1 Simulation Parameters

In order to k-anonymization of the social network graph, the Cuckoo optimization algorithm has been used to add and remove vertices and edges. Table 2 shows the parameters used in this algorithm.

Table 2: Parameters of the Cuckoo Search Algorithm

Parameter	Values	Parameter	Values
Population size	50	Standard deviation	10^{-13}
Maximum number of Cuckoo	25	Accuracy of the answer	0.001
Minimum number of eggs	5	Lower bound	N & min deg
Maximum number of eggs	10	Upper bound	2N & max dog
Egg laying radius coefficient	1	Egg laying radius	0.3
Number of Cuckoo cluster	5	Maximum Iterations	50
-	-	Number of algorithm applying	1

It is very time consuming to set the values attributed to the parameters due to the problem input with big data by trial and error. Given that the problem was solvable by the default parameters of the proposed evolutionary approach, the ratio of population to average error was considered as efficiency criterion. As a result, the smallest population that creates the proper output is selected as the population size of the optimization. Figure 2 shows the process of selecting population size.

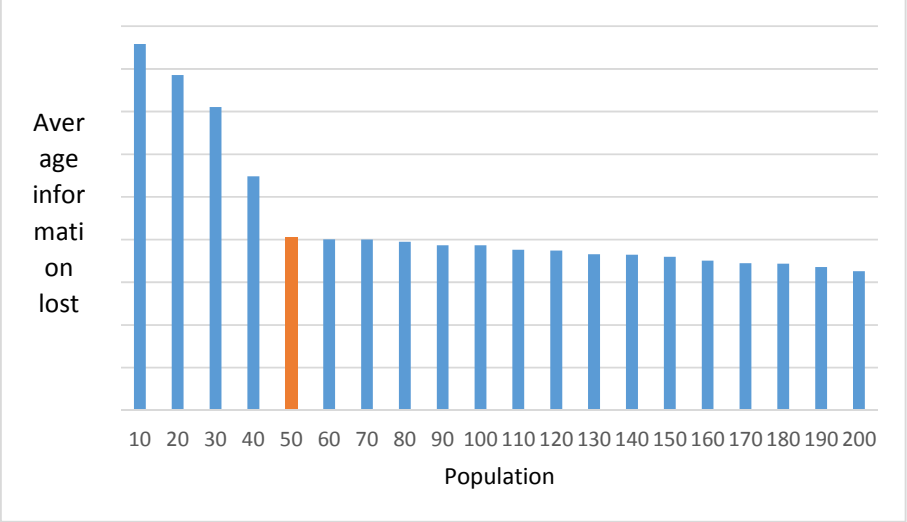

Figure 2: The process of choosing population size based on the average information lost

With regard to Fig. 2, the population of 50 has worse response than more populations, but this amount of difference is more tolerable compared to the overhead entered on the system.

4.2 Results and Comparison

The results of the implementation of the k-anonymity of the social network graph with the combination of clustering methods and Cuckoo's algorithm for the APL, ACC and Transitivity criteria are shown in Table 3. Results have been reported in a different Ks and based on the test standard [13] on the Ca-GrQc dataset.

Table 3: Results of implementing the proposed approach on the Ca-GrQc dataset

Evaluation criteria	K=2	K=5	K=10	K=15	K=20	K=25	K=30	K=35	K=40	K=45	K=50
APL	5.21	5.39	5.44	5.49	5.52	5.59	5.64	5.67	5.74	5.81	5.92
ACC	0.69	0.668	0.68	0.673	0.67	0.667	0.664	0.662	0.66	0.65	0.645
Transitivity	0.68	0.668	0.66	0.656	0.651	0.644	0.633	0.629	0.623	0.61	0.606

The results show that with the complexity of the anonymization process (increase of k), the efficiency of the proposed system appears more due to the effect of the optimization algorithm.

Given that the best results of the Ca-GrQc dataset have been reported by KDVEM and VertexAdd algorithms in reference [13], comparison of the results of the proposed model with these two algorithms has been presented in Figures 3, 4 and 5.

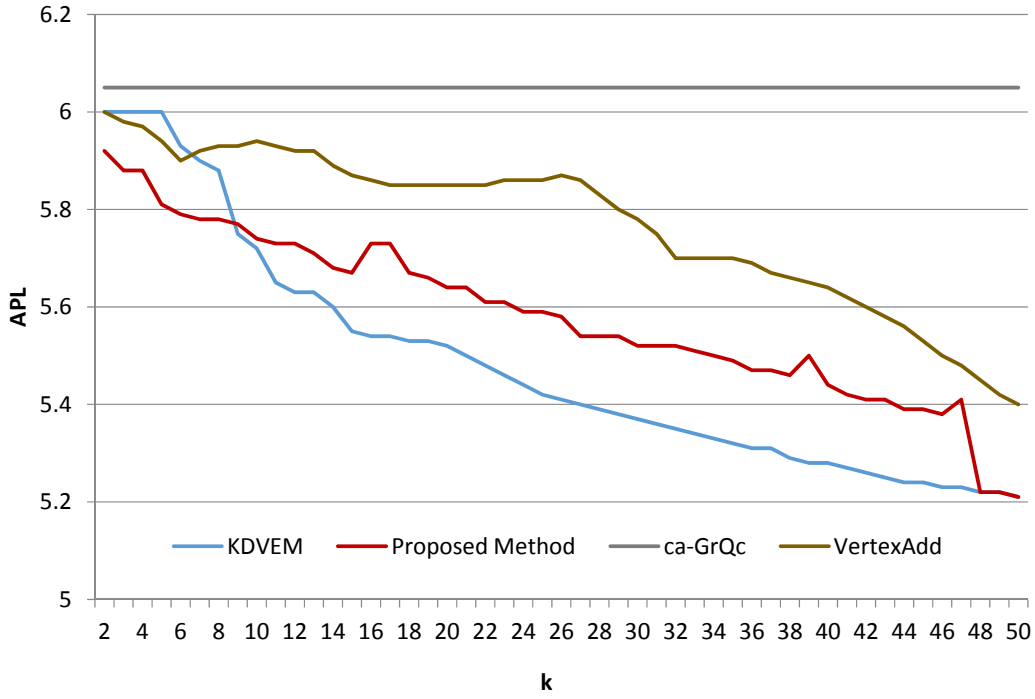

Figure 3: APL results of the proposed model compared to other algorithms

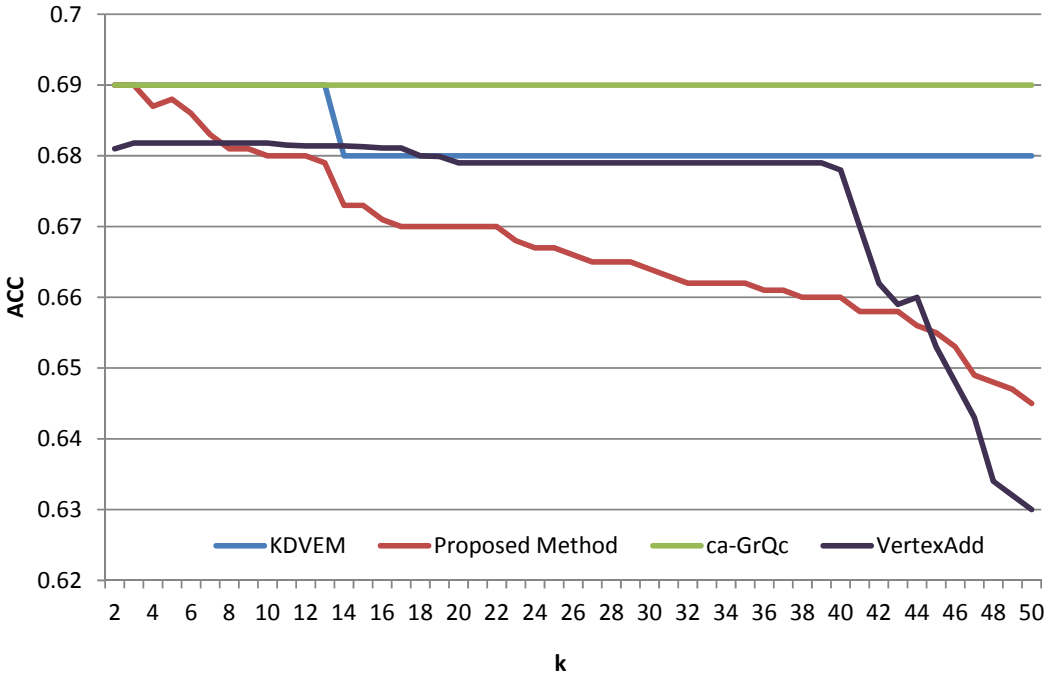

Figure 4: ACC results of the proposed model compared to other algorithms

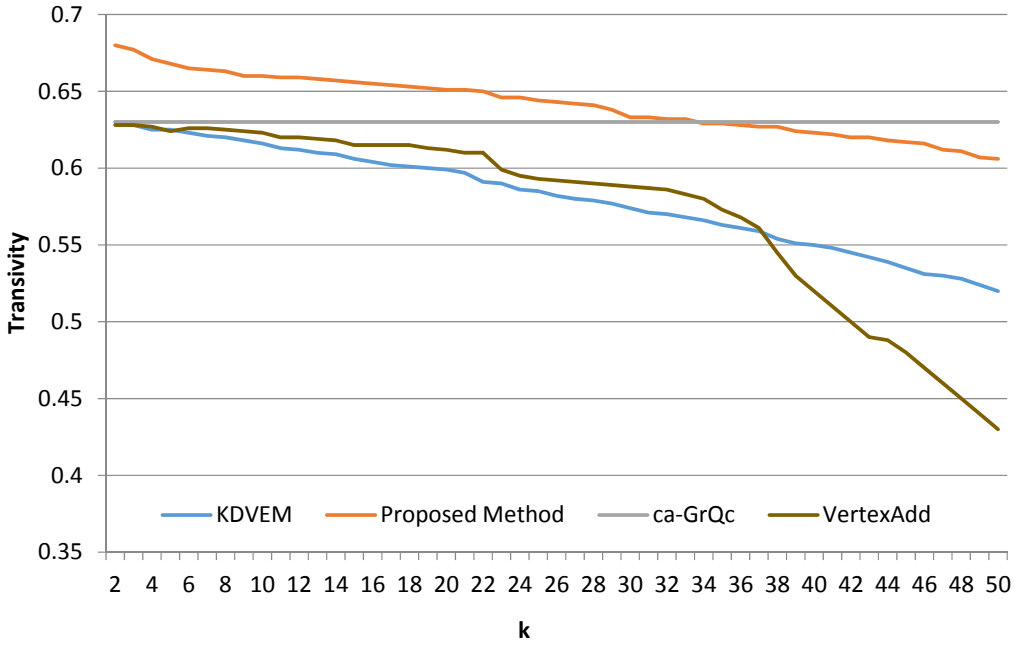

Figure 5: Transitivity results of the proposed model compared to other algorithms

In the k-anonymization, the number of anonymity vertices reaches a maximum value, the amount of anonymity is greater and reduces (optimizing) the parameters of evaluation, which is a better result for anonymity. In fact, the three main parameters of anonymity must be reduced, while the minimize data loss is obtained on output. In Fig. 3, the average results of the suggested path length in smaller Ks, becomes better than KDVEM, and as K becomes larger, the KDVEM gets better results, but generally they are lesser than the main graph and very close to each other. On the other hand, the proposed method has a better result than VertexAdd method. In Figure 4, more anonymity has applied in the average clustering coefficient than the KDVEM method by the proposed method, and better results have been obtained compared with the VertexAdd method, except for some small k at the first and last chart. According to Fig. 5, for the transitivity criteria of k-anonymity result, the KDVEM and VertexAdd methods are better than the proposed method.

Specific arrangement has not yet been proven in terms of priority for the importance of these criteria, but an issue that is important in the anonymity of the social network graph is the reduction of the information lost that is directly related to the assessed criteria. In general, the lower the criterion, the more anonymity, but it leads to increase of the data lost and reduction of the utility. As a result, in another test, the information lost in the proposed method and in the methods of KDVEM and VertexAdd is evaluated compared with the main graph. To calculate the amount of the information lost for different values of k, the change percentage of each one of these three evaluated criteria is calculated compared to its new value, and then their average is considered for the final result. The results of this test are shown in Table 4.

Table 4: Results of the information lost in various methods

Methods	K=2	K=5	K=10	K=15	K=20	K=25	K=30	K=35	K=40	K=45	K=50
KDVEM	1.004	1.19	2.90	4.64	5.11	6.27	7.06	7.91	8.96	10.21	11.37
VertexAdd	1.78	1.34	1.39	2.26	2.66	3.7	4.47	5.9	10.23	15.44	22.69
Proposed method	3.18	3.16	3.57	4.12	4.21	4.33	4.37	4.56	5.28	6.15	6.67

The results show that when k is equal to 10, the proposed method loses more data than the other two methods, but for larger Ks than 10, the proposed method has better results than the KDVEM method and for larger Ks than 25, much better results are reported than both algorithms. Figure 6 shows the average data lost for k-values from 2 to 50 in all three tested methods. The results show that the proposed method is better to maintain the utility of the social network graph in k-anonymization.

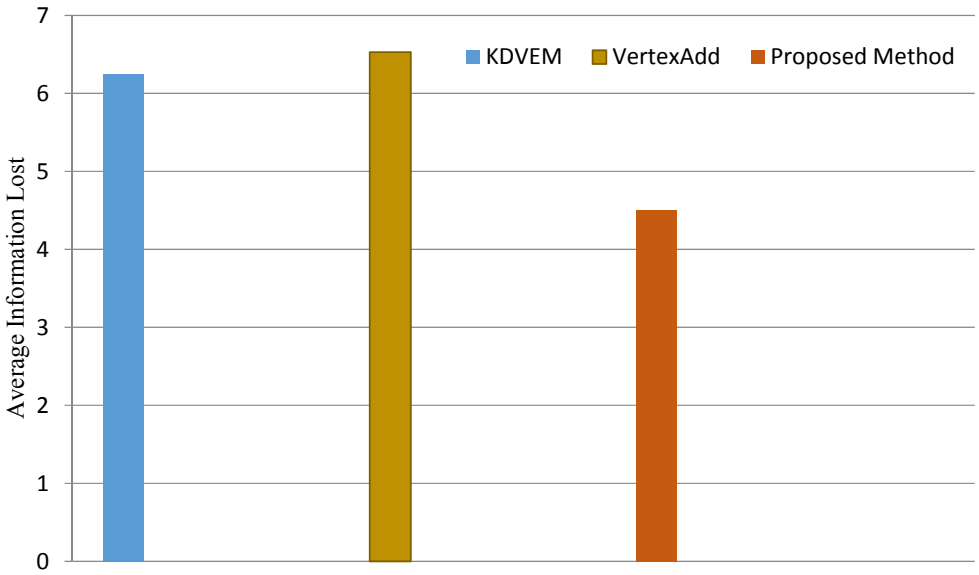

Figure 6: The results of information lost in the proposed model compared to other methods

5. Conclusions and suggestions

By increasing the use of the social networks, and entering the personal information of individuals in these networks, the need for privacy preserving for these networks is felt more than ever. Due to the k-anonymity is one of the most suitable methods for preserving privacy in social networks, a method for k-anonymization was presented in this research using clustering and the Cuckoo optimization algorithm. Preserving privacy through the k-anonymization method has always accompanied distorting the information and thus reducing the utility. For this reason, we must increase the graph's utility by reducing the

information lost. First, data clustering was done in such a way that the least k- vertex with the same degree and lower closeness centrality values are placed in each cluster, then the rest clusters which have less vertex than k and have not yet reached k-anonymity, becomes anonymity by adding vertex and edge. Adding and deleting vertices and edges was optimally carried out by Cuckoo algorithm. The results show that the proposed model has been able to reduce the information lost in the data anonymous and thus it leads to the usefulness of the social network. For future work, clustering based on k- isomorphic or k-adjacency can be used instead of clustering based on k-degree anonymity. It is also suggested that for the k-anonymity clustering, the centrality of the average vertices should be also considered in the prioritization.

References

[1] Musiał K, Kazienko P. Social networks on the internet. World Wide Web. 2013 Jan 1;1-42.

[2] Casas-Roma J, Herrera-Joancomartí J, Torra V. k-Degree anonymity and edge selection: improving data utility in large networks. Knowledge and Information Systems. 2017 Feb 1;50(2):447-74.

[3] Sun Y, Yuan Y, Wang G, Cheng Y. Splitting anonymization: a novel privacy-preserving approach of social network. Knowledge and Information Systems. 2016 Jun 1;47(3):595-623.

[4] Liu X, Xie Q, Wang L. Personalized extended (α, k)-anonymity model for privacy-preserving data publishing. Concurrency and Computation: Practice and Experience. 2017 Mar 25;29(6).

[5] Liu K, Terzi E. Towards identity anonymization on graphs. InProceedings of the 2008 ACM SIGMOD international conference on Management of data 2008 Jun 9 (pp. 93-106). ACM.

[6] Bavelas A. Communication patterns in task-oriented groups. The Journal of the Acoustical Society of America. 1950 Nov;22(6):725-30.

[7] M. Hay, G. Miklau, D.Jensen, P. Weis, and S. Srivastava, "Anonymizing social networks," Science, pp. 1–17, 2007.

[8] Ying X, Wu X. Randomizing social networks: a spectrum preserving approach. InProceedings of the 2008 SIAM International Conference on Data Mining 2008 Apr 24 (pp. 739-750). Society for Industrial and Applied Mathematics.

[9] Boldi P, Bonchi F, Gionis A, Tassa T. Injecting uncertainty in graphs for identity obfuscation. Proceedings of the VLDB Endowment. 2012 Jul 1;5(11):1376-87.

[10] Hartung S, Hoffmann C, Nichterlein A. Improved upper and lower bound heuristics for degree anonymization in social networks. InInternational Symposium on Experimental Algorithms 2014 Jun 29 (pp. 376-387). Springer, Cham.

[11] J. Casas-roma, J. Herrera-Joancomartí, and Torra, V, “Evolutionary Algorithm for Graph Anonymization,” In XII Reunion Espanola sobre Criptologıa y Seguridad de la Informacion (RECSI 2012), pp. 243–248, 2012.

[12] Chester S, Kapron BM, Ramesh G, Srivastava G, Thomo A, Venkatesh S. Why Waldo befriended the dummy? k-Anonymization of social networks with pseudo-nodes. Social Network Analysis and Mining. 2013 Sep 1;3(3):381-99.

[13] Ma T, Zhang Y, Cao J, Shen J, Tang M, Tian Y, Al-Dhelaan A, Al-Rodhaan M. \varvec {\textit {KDVEM}}: a k-degree anonymity with vertex and edge modification algorithm. Computing. 2015 Dec 1;97(12):1165-84.

[14] Zhou B, Pei J. Preserving privacy in social networks against neighborhood attacks. InData Engineering, 2008. ICDE 2008. IEEE 24th International Conference on 2008 Apr 7 (pp. 506-515). IEEE.

[15] Thompson B, Yao D. The union-split algorithm and cluster-based anonymization of social networks. InProceedings of the 4th International Symposium on Information, Computer, and Communications Security 2009 Mar 10 (pp. 218-227). ACM.

[16] Zou L, Chen L, Özsu MT. K-automorphism: A general framework for privacy preserving network publication. Proceedings of the VLDB Endowment. 2009 Aug 1;2(1):946-57.

[17] Wu W, Xiao Y, Wang W, He Z, Wang Z. K-symmetry model for identity anonymization in social networks. InProceedings of the 13th international conference on extending database technology 2010 Mar 22 (pp. 111-122). ACM.

[18] Hay M, Miklau G, Jensen D, Towsley D, Li C. Resisting structural re-identification in anonymized social networks. The VLDB Journal—The International Journal on Very Large Data Bases. 2010 Dec 1;19(6):797-823.

[19] Cheng J, Fu AW, Liu J. K-isomorphism: privacy preserving network publication against structural attacks. InProceedings of the 2010 ACM SIGMOD International Conference on Management of data 2010 Jun 6 (pp. 459-470). ACM.

[20] Tassa T, Cohen DJ. Anonymization of centralized and distributed social networks by sequential clustering. IEEE Transactions on Knowledge and Data Engineering. 2013 Feb;25(2):311-24

[21] Ni S, Xie M, Qian Q. Clustering Based K-anonymity Algorithm for Privacy Preservation. IJ Network Security. 2017;19(6):1062-71.

[22] Chester, S., Kapron, B.M., Ramesh, G. et al. Why Waldo befriended the dummy? k-Anonymization of social networks with pseudo-nodes. Soc. Netw. Anal. Min. (2013) 3: 381

[23] http://snap.stanford.edu/