Securely Publishing Social Network Data

Emad Elabd[1], Hatem AbdulKader[1], and Waleed Ead[2]
[1]Faculty of computers and information, Menoufia University, Egypt
[2]Faculty of Computers and Information, Beni-Suef University, Egypt

Abstract: *Online Social Networks (OSNs) data are published to be used for the purpose of analysis in scientific research. Yet, offering such data in its crude structure raises serious privacy concerns. An adversary may attack the privacy of certain victims easily by collecting local background knowledge about individuals in a social network such as information about its neighbors. The subgraph attack that is based on frequent pattern mining and members' background information may be used to breach the privacy in the published social networks. Most of the current anonymization approaches do not guarantee the privacy preserving of identities from attackers in case of using the frequent pattern mining and background knowledge. In this paper, a secure k-anonymity algorithm that protects published social networks data against subgraph attacks using background information and frequent pattern mining is proposed. The proposed approach has been implemented and tested on real datasets. The experimental results show that the anonymized OSNs can preserve the major characteristics of original OSNs as a tradeoff between privacy and utility.*

Keywords: *Data publishing, privacy preserving, online social networks, background knowledge, anonymization, frequent pattern mining.*

Received May 7, 2016; accepted June 12, 2017

1. Introduction

Online Social Networks (OSNs) such as Facebook and Myspace provide information about individuals in some population and show the links between them. Such links may describe the relations of collaboration, friendship, and correspondence. Some real OSNs are complex and contain a huge set of information. These social networks become an important data source that can be published for different analysis purposes. Therefore, a considerable amount of research has been conducted on social network analysis. For the sake of analysis, OSNs can be modeled as a graph, where the nodes of the graph correspond to the entities and edges denote relations between them. The graph may be modeled directed if the interaction is asymmetric, such as financial transaction network. If the interaction involves more than two parties, the graph can be modeled as hypergraph such as a social network that describes co-membership in social clubs. If there are several types of interactions in the network, the edges can be labeled or the nodes can be accompanied by attributes that provide demographic information such as age, gender, location, or occupation [19].

Many of real-world OSNs contain sensitive information and serious privacy [2, 5, 11, 13]. As a result, research on preserving the privacy of published OSNs data has begun to receive more attention. An important example of a published social network dataset that motivated the study of privacy issues is the Enron corpus. The Federal Energy Regulatory Commission has released a large number of email messages which concern the corporation to the public, of course with the legal related investigation in the accounting fraud and corruption. Such dataset is valuable and available for researchers who are interested in how emails are used for better understanding of organization structure. These data can be modeled as a graph by representing each user as a node and the edge between two nodes means that there is a sufficient email correspondence between such two corresponding individuals. Political blogosphere data is another real dataset example [2]. Such data graph contains over 1000 vertices (nodes) and 15000 edges.

The goal of social network analysis is to uncover hidden social patterns. The range and types of current social network analysis are wider than that of traditional analysis methods which focus on analyzing the attributes of individual social actors. In social network analysis, the relationships and ties between social actors in the network are often regarded more important and informative than the attributes of individual social actors. Social network analysis approaches have been shown very useful in capturing and explaining many real-world phenomena such as the well-known "small world phenomenon" [25]. As a result preserving privacy in publishing OSNs data becomes an important concern.

1.1. Adversary Background Knowledge

Adversaries usually rely on background knowledge to de-anonymize nodes and learn the link relations between de-anonymized individuals from the released anonymized graph. These assumptions of the adversary's background knowledge play a critical role in modeling privacy attacks and developing methods to

protect privacy in social network data. Zhou *et al*. [25] listed several types of background knowledge: attributes of vertices, specific link relationships between some target individuals, vertex degrees, neighborhoods of some target individuals, embedded subgraphs, and graph metrics (e.g., betweenness, closeness, centrality). Some graphs in which nodes are not associated with attributes and links are unlabeled, adversaries only have structural background knowledge in their attacks (e.g., vertex degrees, neighborhoods, embedded subgraphs, and graph metrics). For example, Liu and Terzi [16] considered vertex degrees as background knowledge of the adversaries to breach the privacy of target individuals. Zhou *et al*. [25] usedneighborhood structural information of some target individuals. Zhou *et al*. [25] and Wu *et al*. [21] proposed the use of embedded subgraphs and Ying and Wu [22] exploited the topological similarity/distance to breach the link privacy.

The adversary may use his background knowledge to attack the privacy of some victims by collecting some local knowledge about the target individual vertices in a social network. Consider a synthesized social network of friends as shown in Figure 1. Each vertex in the network represents a person. An edge links two persons who are friends.

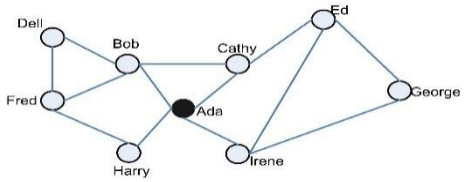


Figure 1. Synthesized social network of friends.

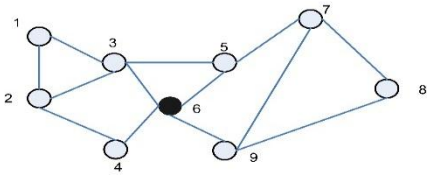


Figure 2. A naïve anonymized social network.

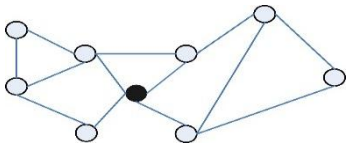


Figure 3. A naïve anonymized social network without node identification.

The SN can be naïvely anonymized by assigning a random ID to each node, e.g., Dell assign to 1, Bob assigns to 2 and so on, as shown in Figure 2. Another way to naïve anonymization the SN by removing the node identification as shown in Figure 3. Suppose that this social network is to be published. To preserve the privacy, it is not sufficient to assign a random ID to each identity; Figure 2; or remove all identities of vertices; Figure 3. If an adversary, unfortunately, has some knowledge about the subgraph of locating an individual, the privacy may still be leaked [25]. The adversary may use his/her backgroundknowledge such as identity subgraph to attack the privacy of some victims.

1.2. Research Problem and Contribution

Many anonymization approaches were proposed to solve privacy preserving of published social networks [1, 4, 8, 12, 15, 23, 25]. Most of these approaches do not guarantee the privacy preserving of identities from attackers in case of using the frequent pattern mining and background knowledge. Graph Frequent patterns are subgraphs that are found from a collection of graphs or a single massive graph with a frequency no less than a user-specified support threshold [3].

The contribution of this paper may be summarized as follow:

1. Developing an anonymization technique to preserve the privacy of released frequent patterns.
2. Providing an effective privacy preserving with a reasonable tradeoff between privacy and information utility of preserving the frequent pattern.
3. Performance evaluation measures such as shortest path length, cluster coefficient and degree distribution [25] have be conducted on real life datasets.

The rest of paper will be organized as follow: Section 2 surveyed most related work in preserving the privacy of released SNs. Section 3 contains the proposed secure k-anonymity framework followed by experimental results and conclusion in Section 4 and 5 respectively.

2. Related work

Fung *et al*. [10] proposed k-anonymity technique to preserve privacy attack based on frequent patterns; one of the most important kinds of knowledge required for marketing and consumer behaviour analysis. Graph frequent patterns are subgraphs that are found from a collection of graphs or a single massive graph with a frequency no less than a user-specified support threshold [13]. In addition, Anusha and Ramana [4] present a framework that provides privacy to individuals in a social network against the adversary from frequent patterns. Their anonymization algorithm is based on the Degree Smoothing method. The first limitation this approach is that it can only deal with 1-neighborhood, if an adversary has the background knowledge beyond 1- neighborhood, the k-anonymous social network may still suffer from neighborhood attacks. The second limitation, it assumed that the adversary has the background knowledge of the structure of the social network. If the adversary has both the structural background knowledge of the social

network and the partial label information of the target victim, this approach is insufficient for this kind of attack. In addition, it doesn't provide information utility measures such as cluster coefficient or shortest path length between social actors.

Social network based trust relationships present a critical foundation for designing trustworthy systems, such as sybildefenses, secure routing, and anonymous/censorship resilient communications. An adversary can reveal the users' trusted social contacts. Liu and Mittal [15] and Chester *et al*. [7] have proposed an anonymization techniques to preserve the trusted user's contacts. The algorithm was based on the use of the degree constrained subgraph satisfaction problem on the complement of the input graph. These algorithms work on anonymizing a given subset of nodes not the entire graph as an adversary with subgraph attack can reveal the others vertices. Moreover, it does not preserve the social frequent pattern after anonymizing the social links.

Abawajy *et al*. [1] present a type of vertex re-identification attack model called neighborhood-pair attack. This attack utilizes the information about the local communities of two connected vertices to identify the target individual. This work cannot protect the relationship (sensitive edge) attack; e.g., k-clique. K-clique is a graph that k- automorphic. Given a k-clique and two individual A and B are in one graph, it's easily to decide that such two individuals are connected by one single edge even though we can't decide which vertex is corresponding to.

Wang *et al*. [20] proposed anonymity algorithm that anonymizes social network data to prevent privacy attacks including both content and structural information. Meanwhile, it preserves the structure information. In the other side, it doesn't preserve the frequent pattern mining knowledge. Zakerzadeh *et al*. [24] prevent the attribute disclosure attack without manipulating the graph structure. In this approach, the pattern of a specified user can easily reveal by an adversary. Song *et al*. [18] and Prashanth and Shaik [17] published their graph in a form such that an adversary who possesses information about a node's neighborhood cannot safely infer its identity and its sensitive labels. The approaches in [9, 17, 18, 24] suffer from preserving a subgraph pattern of an identity from the adversary frequent pattern knowledge. Campan *et al*. [6] proposed k-anonymity to preserve the social network community (subgraph) of a specified node not the entire social network nodes.

Zhou *et al*. [25, 26] show that an adversary may carry out an active attack by maliciously planting some distinct patterns such as connecting sub-graph in a OSNs before it's anonymized and published. They show that k-anonymity and k-automorphism do not guarantee security for Link Information under Neighborhood Attack Graph.

One of the most important kinds of knowledge required for marketing and consumer behavior analysis is the mining of frequent pattern. A graph mining analysis we tend to protect the OSNs individual from locating such pattern (sub-graph) that contains the target.

The previous related works proposed anonymization techniques but it does not guarantee protection to preserve the identity frequent pattern from the attacker's background knowledge. Moreover, it doesn't provide an effective way to preserve the different network properties such as network degree and shortest path lengths. In the proposed algorithm we provide an effective privacy preserving technique that provides with a reasonable tradeoff between privacy and information utility for preserving the frequent pattern.

3. Secure k-anonymity Framework

In this section, we demonstrate the secure k-anonymity framework solution for the anonymized social network (SN) graph.

- *Definition* 1: A social Network (SN) can be represented by a simple graph, G (V, E), where V is a set of vertices and E $\subset$ V x V, is a set of edges. A label function l maps a vertex or an edge to a label. V (G) or E (G) describes the vertex and edge set of G respectively.A graph G' = (V', E') is a subgraph of graph G (V, E), denoted by G' $\subseteq G$, if V' $\subseteq V$ and (u,v) $\in E'$ only if (u,v) $\in E$.
- *Definition* 2: Subgraph isomorphism:For two labeled graphs g and g′, a subgraph isomorphism is an injective function f:V(g) → V (g′), s.t., (1), ∀v ∈V (g), l(v) = l′(f (v)); and (2), ∀ (u, v) ∈E (g), (f (u), f (v)) ∈E(g′) and l(u, v) = l′(f(u), f(v)), where l and l′ are the labeling functions of g and g′, respectively. f is called an embedding of g in g′.

The main idea for solution as follows: Given a SN graph G= (V, E), derive a released graph G_k= (V_k,E_k), Gk is secured K-anonymity, in which G_k={g_1,g_2, g3,…g_k} with pairwise isomorphic $g_i$and g_j, $i \neq j$.

- *Definition* 3: A Secure K-anonymity:Let G= (V, E) be a given with unique node information, for each vertex (node) $v \in V$. Let G_k to be the anonymized graph of G and G_k is secure anonymized with respect to G if for two individual targets vertices A and B with corresponding to the subgraph attack that known by the adversary, the following conditions hold:

For given an anonymized Gk the adversary cannot determine with subgraph pattern belongs to an individual target A. In addition; by frequent pattern two sub graph the adversary cannot determine which pattern belongs to a target A or B. moreover the adversary cannot reveal the sensitive relationship

between A and B by a path with probability of not more than 1/K (secure sensitive edge)

The problem of privacy preserving in publishing a graph by the secure *k*-anonymity is defined as:

- *Definition* 4: Given a SN graph G=(V,E) and positive integer k, such that release an anonymized graph $G_k=(V_k,E_k)$ to be published such that :

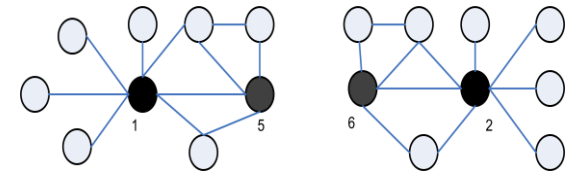
1. $V_k=V$.
2. G_k is secure k-anonymity with respect to G.
3. The anonymization cost from G to G_k is should be minimized.

The proposed solution can be sounded since the published subgraphs g1, g2,…gk are pairwise isomorphic for any subgraph attack for a target individual A. There is at least K different vertices $v_1,\ldots,v_k$ that can be mapped to an individual A and they are not distinguishable vertex identity. Moreover, if the adversary tends to attack the linkage of two individual A and B, in worst case, the adversary can find matched vertices for both A and B in the subgraph g_i and can't distinguish.

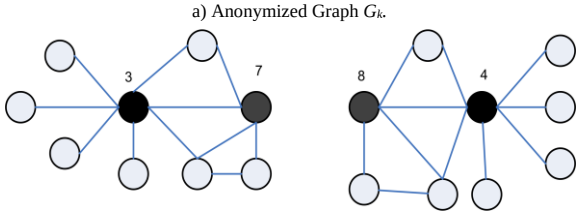
- *Definition* 5: Subgraph Attack**:** an attacker knows the structure of the user and find the structure in the graph or find out which sub graph match with the user structure in the network. In other words, the adversary may know a connected sub-graph G_a, and a vertex V in G_a that may belongs to an individual A.

The secure k- anonymity published graph is based on the secure of anonymized subgraphs, as there are K different vertices a_1, a_2,...a_k that may be mapped to A and k different vertices b_1,b_2,..b_k that can be mapped to B, where $ai \in gi$ and $bi \in gi$ for $1 \le i \le k$.

The following example can demonstrate the above explanation, consider the following Figure 4:



a) Anonymized Graph G_k.



b) Anonymized Graph G_k

c) Sub-graph of A and B.

Figure 4. Published secure *k*-anonymity graph.

Figure 4, shows an example of published secure *k*-anonymity graph. In Figure 4-c, the adversary attack with the two subgraph attacks G_a and $G_b$for the target individuals A and B respectively, which are two nodes in the graph. There are four vertices {1, 2} in Figure 4-a and {3, 4} in Figure 4-b that are linkable to A. While {5, 6} in Figure 4-a and, {7, 8} in Figure 4-b are linkable to B. The adversary can only determine that A and B are linked by an edge with probability ¼. The proposed secure k-anonymity will lead to better privacy preserving of published SN individual's data.

The proposed (Algorithm 1) tends to generate and publish SN graph G_k that consists of identity subgraphs. The set of vertices is preserved by partitioning the *G* graph into *k*-subgraph with the same number of vertices.

In addition, it is difficult to find the appropriate support in a single large graph since multiple embedding of a subgraph may have overlaps. If an arbitrary overlaps between non-identical embedding are allowed, the resulting support will not satisfy the anti-monotonicity property, which is essential for most frequent pattern mining algorithms [3].

- *Definition* 6: Given a pattern p = (V (p),E(p)), a simple overlap of occurrences g and g' of pattern p exists if $g(E(p))\cap g'(E(p)) \neq \emptyset$.

The support of p is defined as the size of the maximum independent set (MIS) of the overlap-graph. A later study [18] proved that the MIS-support is anti-monotone.

- *Definition* 7: Anti-monotonicity means that a size-k subgraph is frequent only if all of its subgraphs are frequent [3]. For each k partition we ensure the each subgraph gi has a symmetry like the others gi's in other k partition by an edge addition or deletion to ensure the Anti-monotonicity of the k-subgraph.

The previous algorithm (Algorithm 1) does not consider frequent of subgraphs. Therefore, Algorithm 2 is proposed as a modification for Algorithm 1 by considering the frequent of subgraphs.

- *Definition* 8: Frequent Graph: Given a labeled graph dataset $D = \{G_1, G_2, \ldots, G_n\}$ and a subgraph g, the supporting graph set of g is $D_g=\{G_i|g\subseteq G_i, G_i\in D\}$. The support of g is support $(g)=|D_g|/|D|$. A frequent graph is a graph whose support is no less than a minimum support threshold, min sup.

The frequent subgraph has been shown to be a sounded strategy in related works such as [8, 25]. In this work, the frequent subgraph has a high possibility to generate large connected subgraphs that minimize the edge modifications needed for the graphs to be isomorphic.

For better performance, we set a threshold on the size of the maximum subgraphs to be considered where the size will be in terms of number of edges in the subgraphs. One way to determine that threshold is the

average degree of G. A justification about determining such threshold is that many vertices in G have this degree d and each form a potential anonymization subgraph with their d 1-neighbors. Such threshold is a suitable basis for locating the frequent subgraphs.

Algorithm 1: K-anonymity Graph

Let K: number of subgraphs,
Vm: vertex mapping for each subgraph gk
Input: graph G and K
Output: anonymized Gk
1. *For each vertex in graph g ,*
Make vertex mapping Vm.
2. *Create gi according to Vm.*
3. *For each gi check isomorphic pairwise by adding or deleting edges.*
4. *Repeat 1-3.*
5. *Return Gk.*

Algorithm 2: Secure K-anonymity Graph

Input: Graph G and an integer K.
Output: an anonymized graph Gk= {$g_1, g_2, \ldots g_k$} of G.
1. *Traverse the given G from each vertex in depth-first manner.*
a. *Enumerate all connected K subgraph.*
b. *If vertex set of these connected subgraphs can't cover all vertices in G*
i. *Enumerate such subgraphs.*
% this means, its isolated components in G%
2. *Extract the vertices of K embedding to be transformed from G to the g_i's in Gk.*
% No necessary to edge modification for anonymization with respect to such embedding %
3. *Remove this embedding from G.*
4. *Anonymize g_i*
a. *Enumerate g's size that is isomorphic subgraphs.*
b. *For each $g' \subset g$ enumerated.*
i. *Search for all embedding of g' in G_k.*
ii. *Such that locate the embedding of these subgraphs within each embedding.*
% Rather than search the big graph G%
iii. *Let T to be a temporary table*
1. *Keep in T every subgraph g' that have been processed a long with embedding embd(g') that have been uncovered.*
% T will help us to check if g' is used or not %
c. *Add edges in each g_i for isomorphic pairwise subgraphs.*
Exit.

For a clarification, we have a non-anonymized G that consists of multiple components (subgraphs), and a *MAX* threshold of number of anonymized subgraphs=*k* and the value of *K*-partition

We enumerate the set of subgraphs with *k* edges. In addition, these subgraphs don't cover the entire graph. Subsequently, do the following:

1. Determine the MIS of such subgraphs.
2. Determine the highest degree in each MIS.
3. Count the number of graphs in each MIS with such degree.
4. Such graph and its MIS are entered into a temporary table T.

4. Experimental Results

For the sake of testing our approach, a real data set is used. This dataset is the EU email[1] communication network. This network was generated using email data from a large European research institution for a period from October 2003 to May 2005 (18 months). We have anonymized information about all incoming and outgoing email of the research institution. For each sent or received email message we know the time, the sender and the recipient of the email. Overall, we have 3,038,531 emails between 287,755 different email addresses. We have a complete email graph for only 1,258 email addresses that come from the research institution. Furthermore, there are 34,203 email addresses that both sent and received email within the span of our dataset. All other email addresses are either non-existing, mistyped or spam. Given a set of email messages, each node corresponds to an email address. We have extracted two datasets (dataset1 and dataset3) from such large dataset with different number of vertices 5000, 10000 respectively to test the proposed approach.

For testing and implementing the proposed algorithm using the different datasets, we use different k values for secure graph anonymization. We use k=5, 10 and generate the anonymized graph for the different datasets. The following Figures 5, 6-a, and 6-b show the degree distribution for the original, anonymized and random graph. The figures show how the proposed algorithm keep and preserve the essential graph information such as degree distribution for different datasets.

Moreover, the following Figures 7 and 8 show transitivity measure which called clustering coefficient; for each node, it measures the proportion of possible neighbor's pairs that are connected. In addition we sample pairs of nodes to find the shortest path length; as shown in Figures 9 and 10. It is observed in both measures transitivity and shortest path length that the proposed algorithm preserves such utility measures.

As previously shown in the experimental results that the released secure k-anonymity graph will preserve the most essential properties such as degree distribution, transitivity distribution and shortest path distribution. As a result, released secure k-anonymity graph can answer aggregate quires with high accuracy. In addition it preserve the privacy against frequent pattern attack as we previously explained.

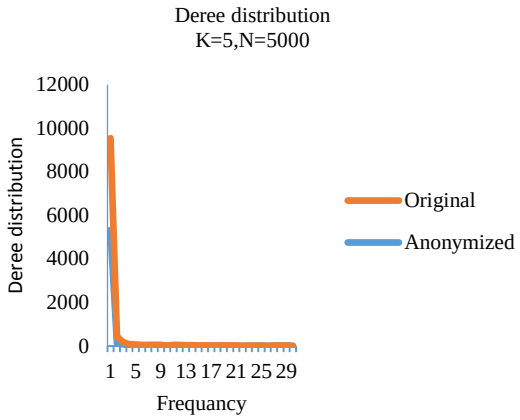
Recall, there is a tradeoff between privacy preservation and its utility [25] as more privacy preserving, more utility loss, and vice versa. As figure 11 shows that the utility extremely loss when k increases as the anonymization cost increases. Subsequently, we should assign the anonymization

[1]http://snap.stanford.edu/data/email-Eu

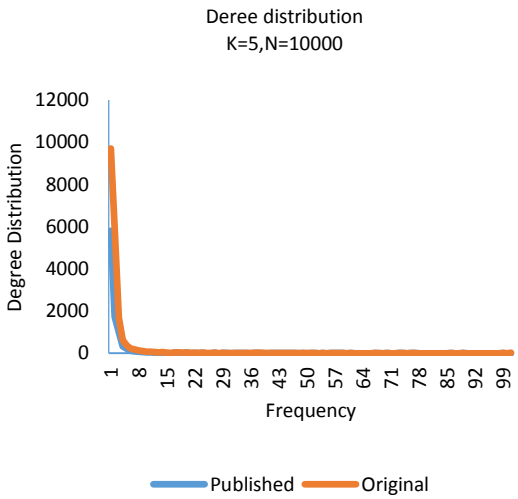
level k such that not to lose the different utility properties to be preserved in the anonymized social network.

5. Conclusions

Publishing social network data is an important issue as these data may contain a treasure of information need to be secured. In this paper, we introduced a secure k-anonymity technique for preventing OSNs identities from adversary's background structural attack. The proposed algorithm has preserved the essential graph information such as degree, transitivity and shortest path distributions against secure graph publishing. As an open issue will be taken in the information associated with each identity and labeled edges and show can such information released securely. Another issue is to propose a privacy preserving framework for publishing the profiles data for OSN's users.

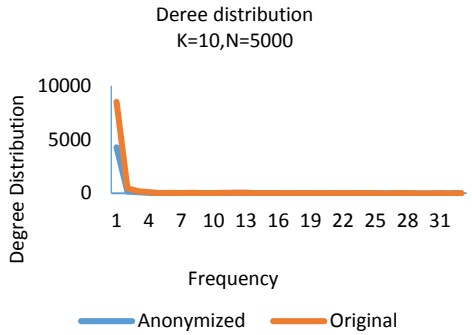


a) Degree distribution for (K=5, N=5000).

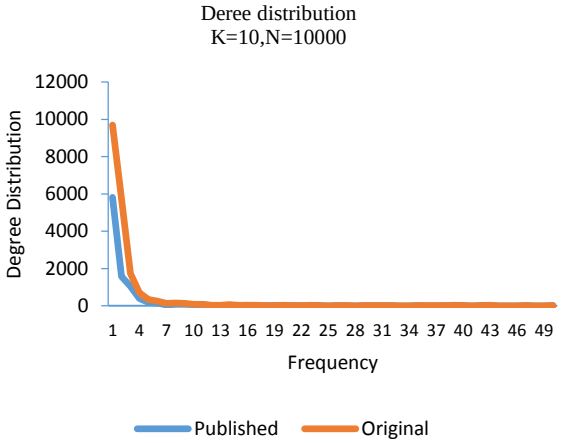


b) degree distribution for (K=5, N=10000).

Figure 5. Degree distribution (k=5).

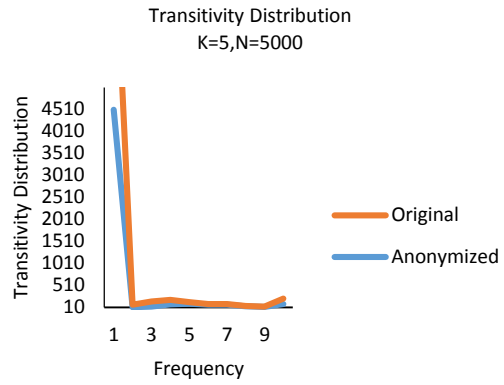


a) degree distribution for (K=10, N=5000).

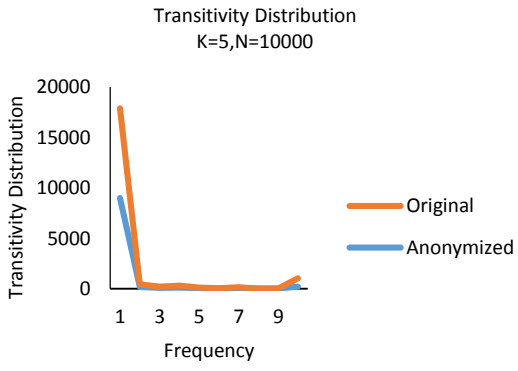


b) Degree distribution for (K=10, N=10000).

Figure 6. Degree distribution (k=10).

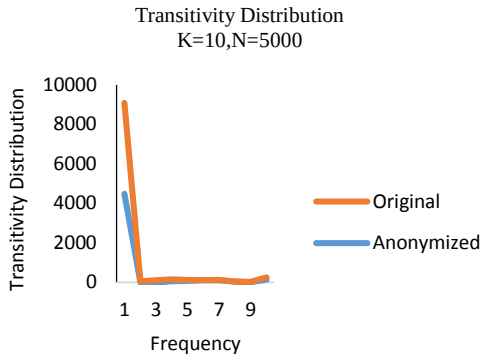


a) Transitivity distribution (k=5, N=5000).

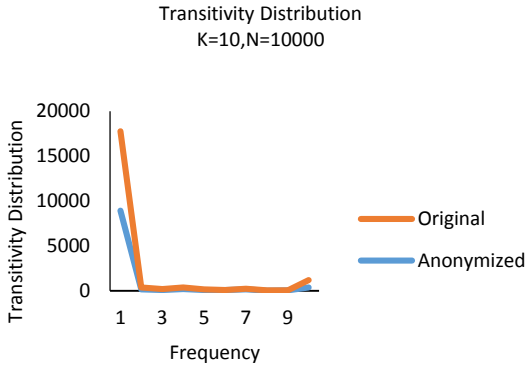


b) transitivity distribution (k=5, N=10000).

Figure 7. Transitivity distribution (k=5).

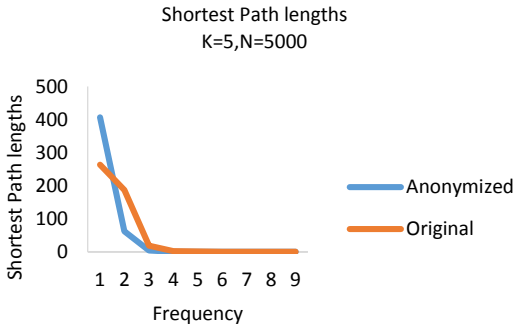


a) Transitivity distribution (k=10, N=5000).

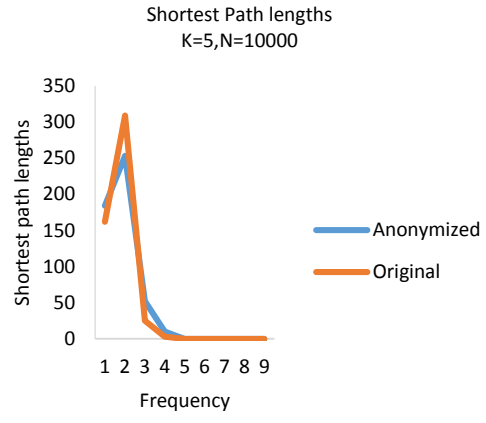


b) Transitivity distribution (k=10, N=10000).

Figure 8. Transitivity distribution (k=10).

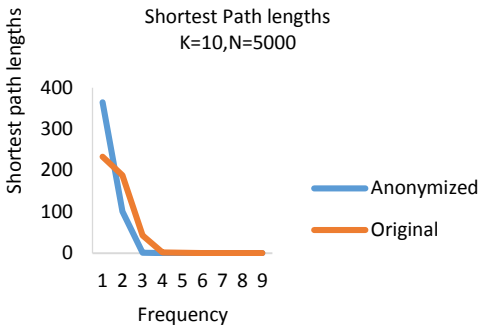


a) Shortest path length distribution (k=5, N= 5000).

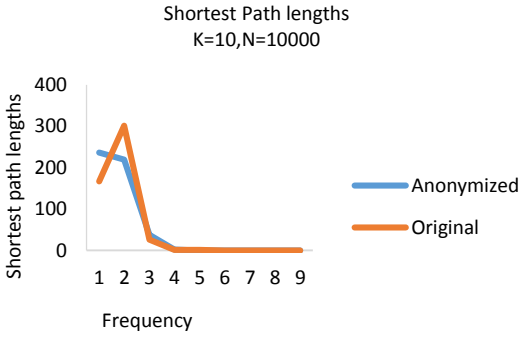


b) Shortest path length distribution (k=5, N=10000).

Figure 9. Shortest path length distribution (k=5).



a) Shortest path length distribution (k=10, N=5000).



b) Shortest path length distribution (k=10, N=10000).

Figure 10. Shortest path length distribution (k=10).

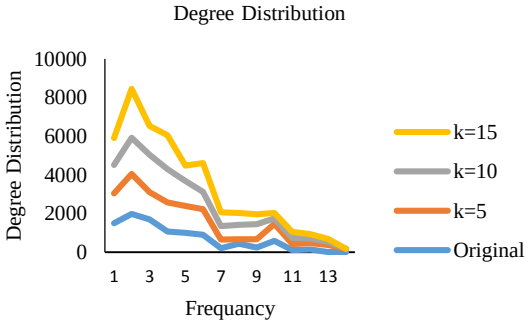


Figure 11. Shortest path length distribution.

References

[1] Abawajy J., Ninggal M., and Herawan T., "Vertex Re-Identification Attack Using Neighbourhood-Pair Properties," *Concurrency and Computation: Practice and Experience*, vol. 28, no. 10, pp. 2906-2919, 2015.

[2] Adamic L. and Glance N., "The Political Blogosphere And The 2004 US Election: Divided They Blog," *in Proceedings of the 3rd International Workshop on Link Discovery*, Chicago, pp. 36-43, 2005.

[3] Aggarwal C., *Mining Graph Data*, Springer, 2015.

[4] Anusha K. and Ramana K., "Degree Smoothing On Social Networks against Frequent Shared Patterns," *International Journal of Advanced Research in Science and Technology*, vol. 4, no. 4, pp. 435-439, 2015.

[5] Backstrom L., Dwork C., and Kleinberg J., "Wherefore Art Thou R3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography," *in Proceedings of the 16th International Conference on World Wide Web*, Banff, pp. 181-190, 2007.

[6] Campan A., Alufaisan Y., Truta T., and Richardson T., "Preserving Communities in Anonymized Social Networks," *Transactions on Data Privacy*, vol. 8, no. 1, pp. 55-58, 2015.

[7] Chester S., Gaertner J., Stege U., and Venkatesh S., "Anonymizing Subsets of Social Networks with Degree Constrained Subgraphs," *in Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Istanbul, pp. 418-422, 2012.

[8] Deshpande A., "A Review on Privacy Priserving Data Publishing of Social Network. Advanced Technologies in Computing and Networking," *in Proceedings of National Conference on Advanced Technologies in Computing and Networking*, pp. 432-434, 2015.

[9] Emelda C. and Jaya R., "Distributed Data Anonymization with Hiding Sensitive Node Labels," *International Journal of Innovative Research in Advanced Engineering*, vol. 1, no. 8, pp. 392-396, 2015.

[10] Fung B., Jin Y., and Li J., "Preserving Privacy and Frequent Sharing Patterns for Social Network Data Publishing," *in Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Niagara Falls, pp. 479-485, 2013.

[11] Hay M., Miklau G., Jensen D., Towsley D., and Weis P., "Resisting Structural Re-Identification in Anonymized Social Networks," *Proceedings of the VLDB Endowment*, vol. 1, no. 1, pp. 102-114, 2008.

[12] Kossinets G., Kleinberg J., and Watts D., "The Structure Of Information Pathways In A Social Communication Network," *in Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, pp. 435-443, 2008.

[13] Leskovec J., Backstrom L., Kumar R., and Tomkins A., "Microscopic Evolution of Social Networks," *in Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, pp. 462-470, 2008.

[14] Li G. and Wang Y., "A Privacy-Preserving Classification Method Based on Singular Value Decomposition," *The International Arab Journal of Information Technology*, vol. 9, no. 6, pp. 529-534, 2012.

[15] Liu C. and Mittal P., "LinkMirage: How to Anonymize Links in Dynamic Social Systems," Technical Report, Cornell Univercity, 2015.

[16] Liu K. and Terzi E., "Towards identity Anonymization on Graphs," *in Proceedings of the ACM SIGMOD International Conference on Management of Data*, Vancouver, pp. 93-106, 2008.

[17] Prashanth R. and Shaik M., "Sensitive Label Privacy Protection on Social Network Data," *IOJETR Transactions on Data Mining*, pp. 1131-1140, 2014.

[18] Song Y., Karras P., Xiao Q., and Bressan S., "Sensitive Label Privacy Protection on Social Network Data," *in Proceedings of International Conference on Scientific and Statistical Database Management*, Berlin, pp. 562-571, 2012.

[19] Tassa T. and Cohen D., "Anonymization of centralized and Distributed Social Networks by Sequential Clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 2, pp. 311-324, 2013.

[20] Wang P., Zhang X., and Huang P., "Privacy Preservation in Social Network Based on Anonymization Techniques," *Computer Modelling and New Technologies*, vol. 18, pp. 249-253, 2014.

[21] Wu X., Ying X., Liu K., and Chen L., "A Survey of Privacy-Preservation of Graphs and Social Networks," *in Proceedings of Managing and Mining Graph Data*, Boston, pp. 421-453, 2010.

[22] Ying X. and Wu X., "on Link Privacy in Randomizing Social Networks," *Knowledge and Information Systems*, vol. 28, no. 3, pp. 645-663, 2011.

[23] Ying X. and Wu X., "Randomizing Social Networks: a Spectrum Preserving Approach," *in Proceedings of the SIAM International Conference on Data Mining*, pp. 739-750, 2008.

[24] Zakerzadeh H., Aggarwal C., and Barker K., "Big Graph Privacy," *in Proceedings of the EDBT/ICDT Joint Conference*, Brussels, pp. 255-262, 2015.

[25] Zhou B., Pei J., and Luk W., "A Brief Survey on Anonymization Techniques for Privacy Preserving Publishing of Social Network Data," *ACM SIGKDD Explorations Newsletter*, vol. 10, no. 2, pp. 12-22, 2008.

[26] Zhou B. and Pei J., "Preserving Privacy in Social Networks Against Neighborhood Attacks," *in Proceedings of the IEEE 24th International Conference on Data Engineering*, Washington, pp. 506-515, 2010.

Emad Elabd Ph.D., Associate Professor, Department of Information Systems, Menoufia University, Egypt. He got his Ph.D. in the field of Web services compliance over high-level specifications at LIRIS, University Lyon1, France, 2011. He received bachelor´s degrees in Electronic Engineering from Menoufia University, Egypt where he did his master's studies in computer science also. His research interests include Web services modeling and analysis with access control and time aspects, Web services (specification, composition), Semantic Web, privacy, LBS, and Information Retrieval.

Hatem AbdullKader obtained hisBSc and MSC degrees, both inelectrical engineering from the Alexandria University, Faculty ofEngineering, Egypt, 1990 and 1995,respectively. He obtained his PhDdegree in electrical engineering alsofrom Faculty of Engineering, Alexandria University,Faculty of Engineering, Egypt in 2001. His areas ofinterest are data security, Web applications andartificial intelligence, and he is specialized in neuralnetworks. He is currently a professor in theInformation Systems Department, Faculty ofComputers and Information, Menoufia University,Egypt, since 2004.

Waleed Ead Lecturer, Information System department, Faculty of Computers and Information, Beni-suef university,Egypt. His BSc and MSc degree in Information system from Zagazig and menoufia university, respectively. He is a Ph.D candidate at menoufia university. His main research interests is privacy preserving in data publishong.