International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I6P134

Volume 72 Issue 6, 380-396, June 2024
© 2024 Seventh Sense Research Group®

Review Article

Comparative Study of Graph-Based Privacy Preservation Techniques

Mariam RAMDI[1*], Oumaima LOUZAR[1], Ouafae BAIDA[1], Abdelouahid LYHYAOUI[1]

[1]*Laboratory of Innovative Technologies (LTI), ENSA of Tangier, Abdelmalek Essaâdi University, Tangier, Morocco.*

**Corresponding Author : mariam.ramdi@etu.uae.ac.ma*

Received: 19 February 2024 Revised: 14 May 2024 Accepted: 29 May 2024 Published: 29 June 2024

Abstract *- Social networks have emerged as subjects of investigation across numerous fields, including sociology, epidemiology, and viral marketing. Analyzing certain structural properties of a social graph, such as node degree or graph diameter, allows for the inference of information about the individuals comprising the network. An effective approach to anonymize a graph involves generalizing specific groups of nodes into supernodes and collapsing multiple links into meta-links. However, it is important to note that this anonymization method may significantly impact the resulting utility derived from the generalized graph. Various research efforts have proposed techniques to anonymize social networks, but the central challenge in this domain lies in achieving a useful final graph with minimal information loss that can be tailored to meet diverse requirements. This article presents a detailed comparative study that elucidates the strengths and weaknesses of different existing techniques found in the literature.*

Keywords *- Anonymization, Social networks, Graph modification, Differential privacy, Generalization.*

1. Introduction

Social networks provide users with convenient access to up-to-date news and information through their advanced functionalities. Prominent social networking platforms like Facebook, LinkedIn, and Google Plus enable users to create profiles and establish connections. These platforms serve a multitude of purposes, including the sharing of thoughts, videos, and images, connecting with friends or seeking new connections, joining groups, and subscribing to preferred communities. While social network data serves as a valuable resource in university research, privacy concerns arise due to the collection and utilization of personal data. The anonymization process emerges as a comprehensive means to safeguard user privacy, transforming data to impede recognition and inference of personal information. However, this process entails a trade-off between privacy protection and data utility.

Numerous scientific research studies have been devoted to examining the preservation of social network users' privacy by proposing anonymization methods aimed at safeguarding their personal information. In [1], the authors described several types of anonymization models that seek to hide or break the link between a real-world person and their sensitive data. Here are the most referenced methods in the literature: k-anonymity [2], t-proximity [3], δ-presence [4], and l-diversity [5]. These methods are supplemented by various relational data anonymization practices and techniques that may be applied to anonymize node attributes in graphs, including generalization [6], random noise [7], character masking [8], data swapping [9], attribute deletion [10], and pseudonymization [11]. Despite these efforts, a thorough comparative analysis of anonymization techniques and their evaluation parameters across multiple highly cited studies is lacking.

This article provides a comprehensive analysis of social network anonymization techniques, categorized into three types: generalization approaches, graph modification, and differential privacy methods. The aim is to identify the most effective method for preserving data utility. Furthermore, a comparative study is conducted to evaluate the anonymization techniques across multiple highly cited studies from 2007 to the present. Subsequently, an effort is made to extract the evaluation parameters used in each study, which measure the data's utility after undergoing the anonymization process.

The findings are summarized, providing valuable insights for future researchers in determining the optimal technique for anonymization procedures. This article seeks to bridge this research gap by embarking on a comprehensive exploration of social network anonymization techniques. These techniques are grouped into three categories: generalization approaches, graph modification, and differential privacy methods. The overarching aim is to pinpoint the most effective method for preserving data utility while simultaneously safeguarding privacy. To accomplish this, a comparative study is conducted across highly cited studies from 2007 to the present. Through this comparative analysis, we extract evaluation parameters used to gauge data utility post-anonymization. The resulting findings not only offer valuable insights for future researchers but also provide a human-centric perspective on the importance of privacy in the digital age.

2. Problem Statement

The generation and collection of personal data have significantly accelerated due to the proliferation of social networking platforms. Traditional anonymization techniques, which generally focus on removing direct identifiers such as names, are inadequate for protecting user privacy against sophisticated re-identification methods. This issue is exacerbated by the complex and interconnected nature of social network data, where relationships between data points pose a challenge in safeguarding private information. Despite the development of several anonymization techniques based on graph modification, generalization, and differential privacy, achieving a comprehensive balance between privacy protection and data utility remains a challenge. Most available research is based on either isolated techniques or specific datasets, lacking a holistic evaluation. By addressing this problem, the study aims to contribute significantly to the field of data privacy and security, offering a thorough understanding of the trade-offs involved in anonymization techniques and guiding the development of more effective methods for protecting user privacy in social networks.

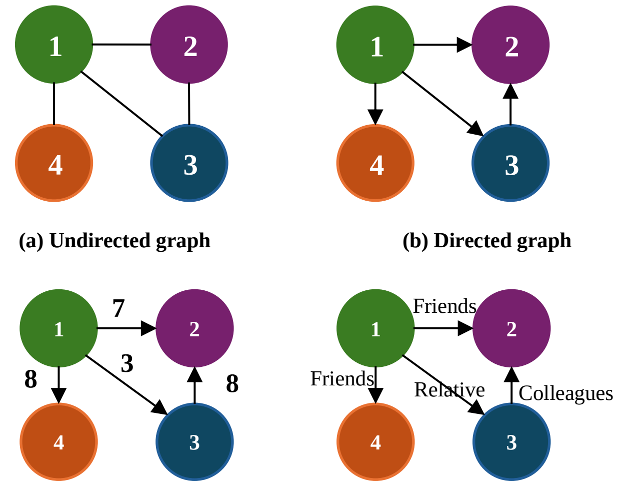
3. Social Networks and Anonymization

The impact of social networking privacy is a crucial consideration in the context of anonymization techniques in social networks. It empowers users by giving them control over their personal information, enhances trust and user confidence, and contributes to online safety and security. Privacy measures also enable individuals to maintain their reputation, promote freedom of expression, and allow for targeted advertising and personalization. Compliance with data protection regulations is also a key aspect. Overall, considering the impact of social networking privacy is essential when exploring anonymization techniques in social networks. Anonymization encompasses various actions and behaviors that enable individuals to maintain their anonymity and withhold personal information, ensuring their rights and privacy are respected. It serves as one of several solutions to utilize personal data while upholding the privacy of individuals. On the contrary, when administrations intend to disclose the data they possess, such as publishing them online, their prior anonymization becomes a legal requirement imposed by the principles governing the relationship between the public and the administration. Consequently, when administrative documents incorporate personal data, they can only be made accessible to the public after undergoing processing, making identification of individuals impossible. Personal data should be anonymized, where appropriate, using methods such as aggregation, pseudonymization, or character masking to reduce the potential risks of privacy breaches, considering the risk of harm associated with the use and non-use of data [12].

4. Representation of a Social Network

A social network can be represented in two different forms: as a graph or as a matrix.Social networks utilize graphs as a means to depict their members and the connections among them. In this representation, the nodes (or vertices) correspond to the social actors, typically individuals within the network, while the links (or edges) signify the relationships between these members. These relationships can be undirected, indicating a symmetrical connection between individuals.

For instance, in the context of Facebook, the "is friends with" relationship is undirected, meaning that if member A is friends with member B, then member B is also friends with member A. In network analysis, an adjacency matrix provides a representation where each vertex in a network is represented by both a row and a column. If two vertices are connected, the corresponding cell at the intersection of the row and the column is marked. Typically, a numerical value is employed, where 0 signifies the absence of a connection, and 1 indicates the presence of a connection. This matrix formulates a concise and structured depiction of the network's connectivity patterns.



(a) Undirected graph **(b) Directed graph**

(c) Directed labeled graph **(d) Directed weighted graph**

V	1	2	3	4
1	1	1	1	1
2	1	1	1	0
3	1	1	1	0
4	1	0	0	1

(e) Matrix sample case to represent social network actors and their relationships.

Fig. 1 Representation of a social network using an Undirected graph (a) Directed graph (b) Directed labeled graph (c) Directed weighted graph (d) matrix (e) with n = 4 nodes and m = 4 links.

5. Expert Knowledge

To develop an effective anonymization process, it is crucial to consider the limitations imposed by the goal of preventing re-identification and the subsequent impact on the data's usability. To construct a relevant anonymization strategy, the following steps are recommended:

- Identify the information that needs to be retained based on its relevance and importance.
- Ensure the exclusion of identifiable information and refrain from including uncommon details that could facilitate the easy recognition of individuals. For example, disclosing the ages of individuals might result in the straightforward identification of centenarians.
- Differentiate between vital information and secondary or non-essential details that can be safely removed.
- Define the desired level of granularity for each stored piece of information, considering both the ideal and acceptable levels.

6. Data Utility

The primary objective of publishing a social graph is to enable analysis of its structure. Evaluating the quality of such analyses is typically achieved through the use of utility measures. In the existing literature, three types of utilities have been extensively considered [13]:

6.1. Structural Properties (Topological)

One of the crucial applications of published social graph data is the analysis of its structural properties. To enhance the understanding and utilization of social graph information, researchers have devised numerous measures to describe the characteristics and structure of a graph from various perspectives. These measures encompass properties such as degree sequence, graph diameter, clustering coefficient, network resilience [14], infectiousness [15], and others. The properties are addressed in studies such as [16], [17], [18], [19], [20], [21], [22].

6.2. Spectral Properties

A graph spectrum refers to the collection of eigenvalues associated with the adjacency matrix that represents a given graph. The spectrum of a graph exhibits a strong correlation with various structural properties inherent to the graph.

6.3. Social Graph Aggregation Queries

An aggregation query performed on a social graph involves the computation of aggregated values for specific properties along certain paths or subgraphs that meet the criteria specified in the query. This type of query, discussed in [23], [24], [25], enables the calculation of consolidated information based on the defined conditions.

7. Types of Attacks

The existing literature on privacy protection in the context of data publication emphasizes the reliance of attackers on contextual knowledge to gain access to sensitive information. These attackers employ various strategies, commonly known as attack models, which involve establishing connections or using probabilistic inferences to deduce sensitive information about their targets.

7.1. Background -Knowledge Attacks

The attacker has sufficient knowledge of the sensitive attribute, allowing him to guess the sensitive data of his victim once he has identified his membership in a group and the age of individuals.

For example, suppose an attacker is aware of the zip code of his victim, and the "patient" table reveals that 52-year-old female patients living in Paris have either heart disease or high blood pressure. If the attacker knows that 80% of active people have hypertension, knowing that their victim is still active, they can conclude that their victim is most likely hypertensive [26].

7.2. Confidentiality Attacks

Attacks on privacy can be classified into three distinct categories, namely active attacks, passive attacks, and semi-passive attacks [27]. These types of attacks are initiated by adversaries who possess a certain level of fundamental knowledge about the target node.

7.2.1. Passive Attack

In a passive attack, the adversary assumes the role of an observer without directly influencing the social network, aiming to comprehend its underlying structure [28]. Passive attacks can manifest in various forms. For instance, in [32], the authors explored the concept of a coalition comprising multiple passive adversaries who act as neighbors in graph G, collaborating to compromise the privacy of other neighboring entities.

7.2.2. Active Attack

In this type of attack, the adversary incorporates a sub-graph and introduces new nodes before anonymization. The attacker then establishes connections between these new accounts and the target nodes. Upon the publication of anonymized data, the adversary utilizes the re-identification of the target nodes and their placements within the social graph to re-identify the embedded sub-graph [30].

7.2.3. Semi-Passive Attack

In this type of attack, no newly created accounts are involved. Instead, the attack involves the creation of links with target nodes before data anonymization.

7.3. Attribute Inference

In [17], the authors demonstrated that it is possible to learn personal information even about users with private profiles through its knowledge of the social graph, thus causing a breach of the privacy of individuals by posting, in most social networks, group memberships of users.

8. System Model

The framework employed in this study is designed to systematically evaluate the effectiveness of various social network anonymization techniques, categorized into graph modification, generalization, and differential privacy methods. The model consists of two primary stages: anonymization and evaluation.

8.1. Anonymization

In the anonymization stage, various techniques are applied to protect privacy within the social network data. These techniques are categorized into three main approaches, each serving distinct purposes (see Section 9 for a detailed explanation)

8.2. Evaluation

The final stage involves assessing the anonymized data using a set of predefined metrics to evaluate both privacy protection and data utility. The evaluation framework draws on parameters extracted from highly cited studies from 2007 to the present, allowing for a comprehensive analysis. The metrics used include:

8.2.1. Privacy Metrics

Measures such as k-anonymity, l-diversity, and t-closeness are used to quantify the level of privacy protection achieved by each anonymization technique.

8.2.2. Utility Metrics

Data utility is evaluated based on various metrics that gauge how well the anonymized data retains its original analytical value. These include information loss, structural similarity, and the accuracy of graph-based queries. (see Section 6 for a detailed explanation).

9. Graph Anonymization Techniques

To preserve the privacy of users within a social graph, various anonymization techniques are employed, involving modifications to the graph structure through the addition and/or removal of nodes and/or edges. These techniques can be broadly categorized as generalization techniques, graph modification techniques, and differential privacy techniques.

9.1. Generalization Techniques

The generalization technique of a social graph consists of grouping several nodes into a single partition called a supernode or cluster and grouping several edges into a super-edge. Three classes of clustering-based techniques are considered. Two algorithms will be presented for anonymizing a social graph by the generalization technique, GraphGen [16], [32] and SaNGreeA.

9.1.1. GraphGen

GraphGen is an anonymization technique that involves partitioning the nodes of a graph into a collection of supernodes. Subsequently, the publication of the number of nodes within each partition, along with the density of existing edges within and across the partitions, is undertaken. The generalized graph is deemed to satisfy the k-anonymity condition if each supernode comprises a minimum of k nodes [16].

9.1.2. Social Network Greedy Anonymization Algorithm (SaNGreeA)

The Social Network Greedy Anonymization (SaNGreeA) algorithm is a technique employed for anonymizing social graphs by generalizing them into clusters or groups of nodes. The objective is to ensure maximum similarity among the nodes within each group. The measure of similarity takes into account both the quasi-identifying attribute values associated with each node and their neighborhood structure [32]. This technique effectively prevents the individualization of a dataset, mitigating the risk of potential correlations with other datasets. Notable generalization techniques in this context include aggregation, k-anonymity [2], l-diversity [3], and t-proximity [5].

9.1.3. K-means Clustering Algorithm

The K-means clustering algorithm is a technique for splitting a data set into k-predefined clusters. It works by minimizing the clustering error, which is the sum of the squared distances between data points and their respective cluster centroids. The algorithm first places cluster centroids randomly and then iteratively changes the position of those centroids by assigning each data point to the nearest centroid and by recalculation of the centroids based on the new memberships of clusters. Although efficient, K-means can be very sensitive to initial placement and can return suboptimal solutions. Usually, multiple runs with different initializations are performed to counter this drawback. The global K-means algorithm can be run incrementally; it adds cluster centers and uses the K-means [33]

9.1.4. Hierarchical Clustering Algorithm

Hierarchical clustering algorithms, through either agglomerative (bottom-up) or divisive (top-down) approaches, organize data into 2a hierarchy of clusters. In agglomerative clustering, all objects are initially in a single cluster, and the algorithm repeatedly merges the nearest clusters until only one cluster remains. 1On the other hand, divisive clustering begins with all objects in a single cluster and splits them recursively into smaller clusters. However, the choice of distance metrics and linkage criteria—like single-link, complete-link, or average-link—plays an important role in shaping the resulting hierarchy. Furthermore, the efficiency can be enhanced by using the nearest neighbor chain algorithm to construct the hierarchical clustering more efficiently than classical algorithms. [34]

9.2. Differential Privacy

This anonymization technique involves query anonymization through the introduction of noise. Rather than directly anonymizing a database, it perturbs the results of a numeric query executed on certain statistical properties, such as the average of a variable or the frequency of a modality. One of its key advantages is the provision of statistical guarantees regarding the level of confidentiality it offers. Differential privacy stands as the sole method that provides formal guarantees and mathematical proofs of the ability to restrict the information that can be inferred about individuals. This method incorporates both the sampling of actual data (with probability α) and the generation of synthetic data (with probability β << α) while ensuring the synthetic data remains realistic. Another notable strength of differential privacy is its adaptability. Unlike the conventional practice of sharing raw data, the results generated through differential privacy can be tailored on a case-by-case basis, taking into consideration the specific requests and the authorized third parties involved.

This adaptive approach facilitates effective governance and addresses concerns related to data sharing. Formal guarantees are crucial and make it possible to quantify the possibility of re-identification of tuples, hence the enthusiasm for this method. Indeed, by observing the anonymous dataset, the information that can be obtained on whether a tuple is true or false is doubly bounded: it is never certain that a tuple is true with a probability greater than α, nor that it is false with a probability less than β.

Differential privacy techniques are classified as node privacy techniques [35], [36], and edge privacy techniques [37]. In conclusion, one of the limitations of this technique is its restriction on sharing the dataset in its original structure, thereby limiting the scope of feasible analyses.

9.3. Graph Modification

9.3.1. Randomization

This method involves a two-step process, namely edge removal and edge addition. Initially, the data owner selects a parameter, k, and removes a randomly selected subset of edges from the graph. In the subsequent step, the owner introduces new relationships by randomly connecting k pairs of previously unconnected nodes. The introduction of random perturbations utilizes a Bernoulli sequence of draws. In their study [38], the authors have introduced two distinct notions of privacy protection. The first notion aims to safeguard against adversaries attempting to identify a particular individual within the perturbed graph, while the second notion focuses on protecting against adversaries without specific target individuals. Numerous works have been proposed in the literature employing the randomization technique, as evidenced by [39], [40], [41], [42].

9.3.2. K-anonymity

The underlying principle of k-anonymity is to obfuscate the association between individuals and their corresponding records by blending individuals within a larger group. A table adheres to the concept of k-anonymity if each record in the table cannot be distinguished from at least k - 1 other records.

9.3.3. K-degree Anonymity

The concept of k-degree anonymization ensures that for every node v ∈ V in a graph G = (V, E), there exist at least k-1 other vertices in the graph that have an identical degree as v. Mathematically, this can be represented as: In [44] [45], the authors opted to introduce false vertices rather than a set of edges. The core idea of their approach involves establishing connections between the original vertices and the artificially generated ones.

9.3.4. Degree-Based Anonymization Technique

Typically, the node degree is regarded as a primary characteristic, and adversaries leverage this property as fundamental knowledge in degree-based approaches to identify the central vertex. For instance, if an adversary possesses information about a vertex that is connected to a specific number of neighbors, they can potentially identify the targeted node if there exists only one node with that precise number of neighbors.

9.3.5. Neighborhood-Based Anonymization Techniques

The fundamental concept behind this technique is that the attacker utilizes the knowledge derived from the immediate neighbors of a node. Therefore, the neighborhood of a node v ∈ V in a graph G represents a subgraph consisting of the neighboring vertices of the original vertex v [23], [24].

9.3.6. K-neighborhood Anonymity

In this anonymization technique, for each node $\nu \in V$ in a graph G = (V, E), the node is considered k-anonymous if there exist at least k - 1 other nodes such that (ν_1), $N(\nu_2)$, ..., $N(\nu_{k-1})$ are isomorphic, where $N(\nu_i)$ represents the adjacent subgraph of node ν_i. According to the authors in [46], the attacker possesses knowledge of the common neighbors between two directly connected nodes. They propose that for every edge e ∈ E, there should be at least k - 1 other edges that share the same number of common neighbors as e.

9.3.7. Subgraph-Based Anonymization Technique

In this technique, the adversary regards the subgraph as fundamental information for identifying the target within the graph. This knowledge is represented as a query that yields a subgraph as its result.

9.3.8. K-automorphism

A graph G' (V', E') is considered k-automorphic if, for each node ν, there are at least k - 1 automorphic functions that exist. To effectively preserve the utility of the published graph, the algorithm needs to minimize the inclusion of false edges.

This requirement implies that the subgraphs within a group Ui should exhibit a high degree of similarity to each other, thereby ensuring that step 2 introduces only a small number of edges. Additionally, the presence of a limited number of edges across different subgraphs is desired to prevent step 3 from adding a significant number of edges [23].

10. Related Work and Discussion

The preservation of privacy for users of social networks has emerged as a prominent research area. Anonymization has emerged as a solution to address this challenge. An effective anonymization process must take into account two key aspects. Firstly, it should prioritize the preservation of the confidentiality of individuals within the social graph. Secondly, it should strive to maintain the utility of the data even after anonymization, ensuring that the anonymized data remains valuable for various purposes, including scientific research. In this section, the research conducted in this domain will be categorized into three types of techniques: generalization techniques, graph modification techniques, and differential privacy techniques.

10.1. Generalization Research

In this technique, the anonymized graph exhibits an increase in both the total number of nodes and edges. In a study conducted by the authors [47], a novel approach was proposed for anonymization that involved generalizing the graph structure and partitioning the nodes based on an analysis of the social graph at the partition level. The significance of incorporating a wide margin of analysis in the final stage of anonymization was emphasized to mitigate the risk of re-identification and enhance the overall protection. The authors in [48] place particular emphasis on maintaining two crucial aspects throughout the anonymization process: algorithm efficiency and data utility preservation. Their proposed algorithm comprises three distinct steps. The first step involves globally introducing noise to the vertices of the graph. The second step aims to ensure that the amount of added noise remains within acceptable bounds.

Finally, the third step validates the proposed approach through its application in various tasks such as clustering, classification, and node categorization. The authors in [49] have introduced an algorithm that effectively addresses two key criteria: minimizing the loss of structural information and achieving effective generalization. Their method involves bucketizing the predefined attribute variables of the vertices based on the similarity of values, thereby creating classes. These classes are then evaluated by calculating the average information loss locally for each class. Finally, the algorithm creates super nodes by selecting the best class obtained from the previous step. This approach ensures a balance between preserving structural information and achieving effective generalization.

10.2. Differential Privacy Research

Anonymization techniques employing differential privacy methods incorporate the fundamental concept of introducing random noise into the original data set. This includes actions such as adding, removing, or randomly altering edges between nodes in the social graph. Numerous anonymization methods have been proposed in this domain, including works by researchers [35], [15], [17], [50], and [51]. These methods contribute to the advancement of privacy-preserving techniques by leveraging the principles of differential privacy and its application to social network data. In their study [52], the authors opted for a combination of three series, namely dk-1, dk-2, and dk-3, to achieve their primary objective of maximizing the retention of data utility following the anonymization process. By leveraging these three series, their approach aims to strike a balance between privacy protection and preserving the usefulness of the anonymized data.

The authors in [53] introduced a novel anonymization method that focuses on local noise addition within the social graph. In their study, they conducted a comparative analysis of their proposed method against global differential privacy techniques. By evaluating the effectiveness and utility preservation of their approach, they aimed to provide insights into the advantages and limitations of local noise addition for privacy protection in comparison to global approaches. In their work [54], the authors introduced an anonymization method that leverages local differential privacy techniques. Their approach involved utilizing Hierarchical Random Graph (HRG) methods to preserve differential privacy guarantees while reducing noise locally within subsets of the social graph. By adopting this strategy, they aimed to strike a balance between privacy protection and maintaining data utility at a more granular level within the graph structure.

10.3. Graph Modification Research

There exist numerous techniques for anonymizing user identities, with one of the pioneering methods being the introduction of k-anonymity by Sweeney [2]. This approach was devised to safeguard against identity disclosure by anonymizing information, thereby mitigating structural attacks targeting node degree identification within social network graphs. To adapt this method specifically for social network graphs, a variant known as k-degree anonymity was proposed [16]. This approach employs the concept of k-anonymity at the vertex level, ensuring that each vertex in the graph has at least k-1 other vertices with the same degree. To effectively implement k-degree anonymity, several innovative methodologies have been put forth, exploring novel techniques and algorithms.

The authors in [3] conducted a comprehensive investigation into the speed and scalability aspects of the k-degree anonymity method using various heuristic techniques. As a result, they proposed an enhanced approach that combines the addition of edges with the anonymization of vertex degree sequences in groups. Building upon this method, in [55], the author further advanced the state-of-the-art in their work. In their work [56], the authors introduced a novel anonymization method focusing on adding noise to nodes to achieve k-degree anonymity. They also devised an effective grouping strategy for the social network's servers, which involved generating and distributing information among them during the anonymization process.

In their publication [57], the authors have presented a method that addresses the criteria of preserving data usefulness in the anonymized graph, thus ensuring a high quality of anonymization.

10.4. Discussion

It can be concluded that there is not a definite "best" method among the techniques recently mentioned. The choice depends on the specific need, the data, the level of privacy desired, and the trade-offs one is willing to make between privacy and the usefulness or accuracy of the data. Graph modification, generalization, and differential privacy represent the three most prominent approaches to anonymizing social network data, each with its distinct methodological characteristics and implications for privacy and utility. Graph modification techniques, essentially ways to alter the structure of a network to veil sensitive relationships, tend to distort the topology of the network, sometimes considerably, therefore reducing analytic utility and increasing the likelihood of overfitting to attack scenarios. Generalization reduces data to higher-order categories, thus simplifying the data but usually entailing significant information loss. The loss reduces the granular resolution of data, which reduces its usefulness for any detailed analysis and possibly makes it miss subtle patterns that are critical for some applications. Differential privacy, in contrast, is an extremely mathematically rigorous and versatile framework that achieves strong privacy guarantees while still ensuring that data utility is preserved. In contrast to graph modification, differential privacy perturbs data through added noise in a manner that affects minimal, if any, alteration to their statistical properties. This also helps to avoid the huge information loss related to generalization.

Besides, the formal definition of privacy in differential privacy makes the approach resistant to several re-identification attacks; the protection is quantifiable and can be adapted to very diverse data types and analytical tasks. In conclusion, differential privacy offers a better balance between privacy and utility compared with graph modification and generalization. It offers the best guarantees for its robust, formal privacy that is flexible and does not affect the usability of data.

11. Comparison

Table 1 provides an overview of various anonymization techniques that have been proposed to address the issue of confidentiality. These techniques are categorized into three main categories: generalization, graph modification, and differential privacy, with an additional category of machine learning classification. Many of these methods serve as foundational approaches for the development of more recent anonymization techniques in this field.

Anonymization is a technique aimed at mitigating the risk of re-identification by introducing complexity and cost to such attempts. Nevertheless, with the advancement of data mining technologies and the proliferation of large-scale open databases, anonymized data is becoming increasingly vulnerable to attacks.

Indeed, an anonymized dataset is inherently vulnerable to re-identification attacks, as it remains susceptible to linking and cross-referencing with other datasets to uncover individuals' identities and reveal sensitive personal information.

Table 1. Related work on anonymization techniques

Anonymization Method	Ref	Year/ Citations	Anonymization technique	Input data type	Data	Input	Output
Differential privacy (random alteration)	[57]	2008/360	Spectrum randomization	• Directed graph • Adjacency matrix	• US politics book data	• Original graph	• Spectrum preserving randomized graph
	[17]	2007/412	Random perturbation	• Undirected, unlabeled graph	• Hep-Th • Enron • Net- trace • Net- common	• Original graph	• Randomly perturbed graph
	[43]	2009/280	Graph degree distribution		• Flicker • Orkut • LiveJournal • YouTube	• Original graph	• Perturbed graph
	[58]	2009/77	Edge randomization		• Polbooks • Polblogs • Enron	• Original graph	• Perturbed graph
Generalization	[59]	2016/251	Attributes Generalization	• Undirected, labeled graph	• Facebook	• V, E, X(attributes), Yk(labels of known users) • Core, Ɛ (utility	• Yu (labels of unknown users) • Anonymized Graph

						threshold)	
	[60]	2017/94	Weight Generalization		• Facebook • CA-CondMat • Enron • Douban	• G(u), G(v), DF (different damping factors) • Graph Groups	• Cost (Ge(u), Ge(v)) • Anonymized Graph
	[14]	2008/620	Partitioning the nodes and summarizing the graph at the partition level	• Undirected, unlabeled graph	• Hep-Th • Enron • Net-trace	• Original graph	• Generalized graph
	[56]	2010/111	Nodes Generalization		• Tree • HepTh • HOT • Enron • Mesh • NetTrace • Power-Law	• Original graph, minimum supernode size	• Generalized graph
Graph modification	[16]	2008/860	K-anonymity: anonymizing vertices		• Enron • Prefuse • Powergrid • Scale-free • Smallworld • Random graphs	• Original graph	• K-degree anonymous graph. Dynamic programming.
	[16]	2012/53	K-degree: adding edges and anonymizing its degree sequence		• Enron	• v: Sorted vertices by degree, i: an index, k: the value of anonymity	• K-degree anonymous graph
	[61]	2015/88	K-degree anonymity (vertex and edge modification)		• Ca-HepTh • Enron • Ca-GrQc • Ca-AstroPh • Ca-CondMat	• Original degree sequence d, anonymization level k • G, k, d, d', n	• K-anonymous degree sequence d' • Anonymized Graph
	[14]	2008/824	K-anonymity adds fake edges	• Undirected, labeled graph	• High Energy Physics	• Original graph	• K-neighbourhood anonymous graph.

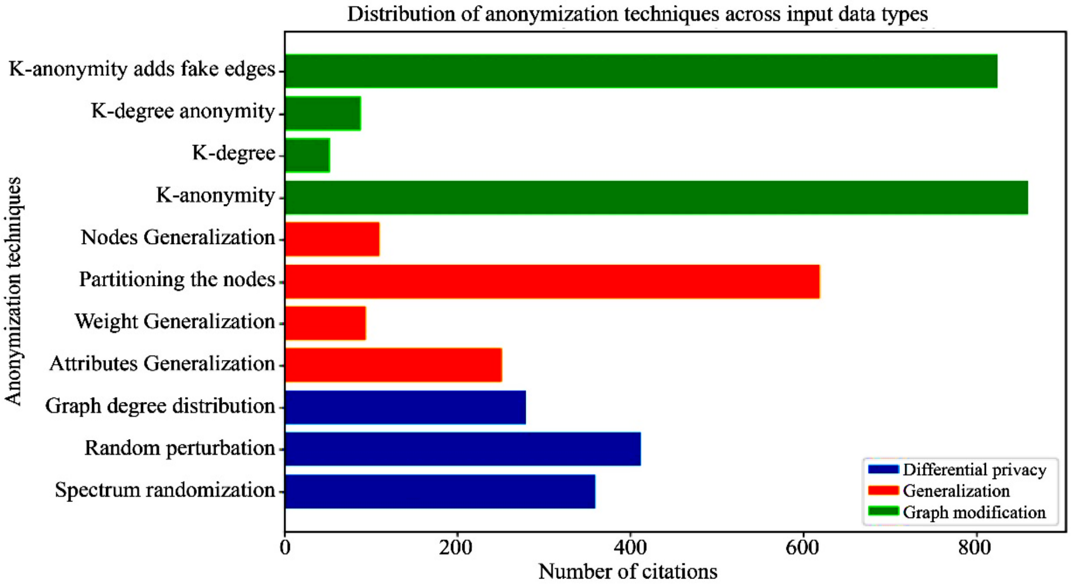


Fig. 2 Distribution of anonymization techniques across input data types

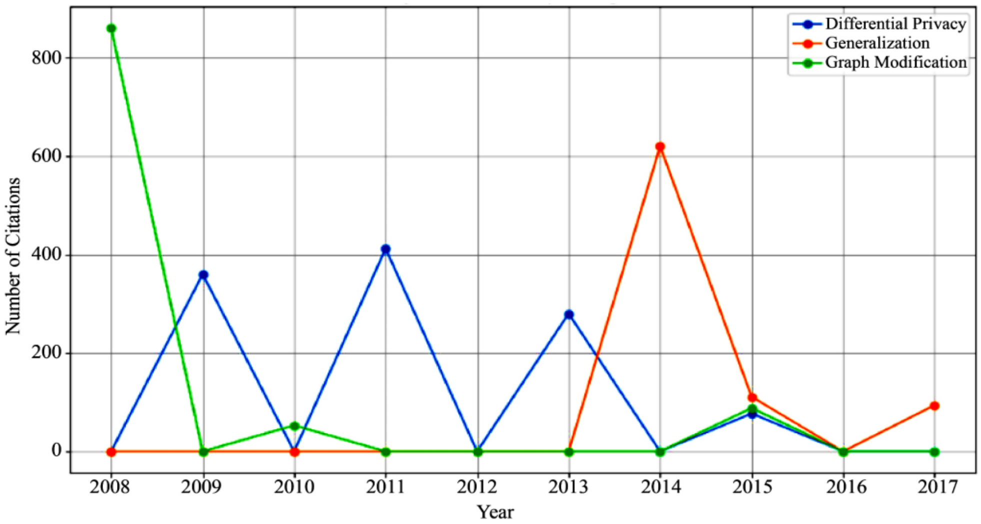


Fig. 3 Trend of anonymization technique over the years

Table 2. Abbreviation

Abbreviations	Full Name
DP	Differential privacy
K-DA	K-Degree anonymity
RW	Random walk
Deg.	Degree
ED	Effective diameter
APL	Average path length
CCoef	Clustering coefficient
CC	Closeness centrality
BC	Betweenness centrality
SE	Structural entropy
EN	Number of edges

Generally, achieving optimal anonymization requires striking a balance between two key factors:

- The risk of re-identification of individuals, which arises from the potential linking of partial information fields across datasets.
- The risk of information loss, ensuring that the anonymized data remain relevant and useful for their intended purposes.

The authors employ various parameters, as described in Section 5, to assess the utility of data following the anonymization process. In this section, the aim is to comprehensively list and evaluate each of the evaluation parameters utilized in numerous studies.

Table 3. Analysis of the effect of anonymization on graph measurements on Enron data.

Ref	Enron						
[15]	**Measure**	**Deg.**	**ED**	**APL**	**CC**	**BC**	**CCoef**
	Original	5.0	9.0	4.0	0.276	0.005	0.286
	Perturbed5%	4.5	8.7	3.2	0.293	0.009	0.242
	Perturbed10%	4.6	7.6	3.0	0.304	0.010	0.191
	Perturbed100%	5.0	6.1	3.0	0.337	0.014	0.000

Table 4. Average path length and Clustering coefficient for the spectrum of Greedy Swap graphs.

Ref		APL				CCoef
		Original	**Super Graph**	**Priority**	**Greedy Swap**	**Greedy Swap**
[16]	α = 2	~29	~7	~8	~8	~0.70
	α = 4	~25	~11	~7	~9	~0.79
	α = 6	~5	~5	~5	~5	~0.78
	α = 8	~3.2	~3.2	~3.2	~3.2	~0.42
	α = 10	~2.5	~2.5	~2.5	~2.5	~0.1
	α = 12	~2.5	~2.5	~2.5	~2.5	~0.02
	α = 14	~2.5	~2.5	~2.5	~2.5	~0.01

Table 5. Average path length, effective diameter, betweenness centrality and radius measures for different k.

Ref			ED	APL	Radius
[56]	Original		9	2.9517	5
	Anonymous	K=0	9	2.9517	5
		K=5	9	2.9895	5
		K=10	9	2.9506	5
		K=15	9	2.9302	5
		K=20	9	2.9123	5

Table 6. Data utility and privacy measurement under link removal, attribute removal and collective methods.

Ref	Dataset	Utility/Privacy		
		Collective	Attribute Removal	Link Removal
[48]	SNAP	1.1967	1.5273	1.2636
	Caltech	1.1639	1.3433	1.1881
	MIT	1.1639	1.3433	1.1931

It should be noted that all values denoted by the tilde (~) in this section have been derived from graph data and are therefore treated as approximate values. Initially, the intention was to consolidate all the research findings into a single table. However, a challenge was encountered due to the variations in evaluation parameters chosen by different authors, which directly reflect the effectiveness and outcomes of the respective methods employed. Consequently, it was opted to present each research study individually in this section. In [15], the authors employed an add/remove and switch method to modify the edges in the social graph globally, resulting in changes to its structural properties. This approach effectively preserves the usefulness of the data in the conditionally anonymized graph. They evaluated their method using various parameters, including Deg, Diam, APL, CL, BT, and CCL. Table 4 presents the calculated parameter values that measure the utility of the data after applying their method to the Enron dataset. In their work [16], the authors employed a k-anonymization technique to anonymize the data in the graph. They selected the APL and CCoef parameters to evaluate the effectiveness of their approach. The results of their study indicate that the proposed anonymization process does not significantly obscure the inherent properties of the original graph.

The authors in [56] employed the evaluation metrics of ED, APL, and Radius to assess the performance of their algorithm. The results presented in Table 5 demonstrate that their method yields negligible changes in the attributes of the input graph. Notably, the attributes only vary when the value of k is altered during program execution. Based on their findings, the authors concluded that the algorithm produces satisfactory outcomes when there is minimal alteration in the characteristics of the social graph. Table 6 presents the maximum utility/privacy achieved through the application of attribute removal, link removal, and collective methods during the anonymization process on SNAP, Caltech, and MIT datasets [48].

In their study [49], the authors employed Deg, APL, and CCoef as evaluation metrics. Table 6 presents the results obtained after comparing the utility of the data following the execution of the anonymization algorithm on the Facebook and Enron datasets. The authors observed that increasing the value of the parameter k leads to a decrease in the utility of the data in the anonymized graph.

Table 8 presents the impact of anonymization on the clustering coefficient. In their study [14], the authors observed that applying their anonymization algorithm to the HepTH and Enron graphs resulted in a decrease in the clustering coefficient, indicating a reduction in the graph's clustering structure.

Table 7. Utility of the anonymized social network.

Ref			Deg	APL	CCoef
[49]	Facebook	Original	44	4.7	0.605
		K=5	45	4.06	0.537
		K=10	47	3.98	0.508
		K=15	49	3.85	0.478
		K=20	52	3.81	0.468
		K=25	54	3.73	0.466
	Enron	Original	10	4.9	0.497
		K=5	10.4	4.92	0.496
		K=10	10.6	4.99	0.4813
		K=15	11	5	0.476
		K=20	11.2	5.01	0.445
		K=25	11.3	5.05	0.423

Table 8. Shortest-path length distribution and Clustering coefficient of Enron and HepTh graphs.

Ref			Path length		Clust. Coeff			
					Original	GraphGen	LT	BCKS
[14]	HepTh	K=2			~0.25	~20.18	~0.09	~0.06
		K=5			~0.25	~0.12	~0.07	~0.00
		K=10	H1	1.369	~0.25	~0.07	~0.06	~0.00
			H2	0.947				
		K=20			~0.25	~0.03	~0.07	~0.00
	Enron	K=2				~0.32	~0.28	~0.10
		K=5				~0.27	~0.24	~0.05
		K=10	H1	6.174		~0.22	~0.23	~0.06
			H2	0.793				
		K=20				~0.15	~0.20	~0.06

Table 9. Relative performance of different algorithms by metric.

Ref		Transitivity	ACC	APL	Average
[63]	Priority	3.277	3.615	3.631	3.508
	KDVEM	2.046	1.631	2.554	2.077
	FKDA	1.985	2.046	2.277	2.103
	Vertex-Add	3.508	2.077	2.103	2.313

Table 9 provides a comparison of various algorithms based on different metrics. In their study [31], the authors introduced a method called KDVEM, which demonstrates satisfactory performance in terms of the ACC metric. Additionally, the FKDA method outperforms other algorithms in terms of the transitivity metric, while the vertex-add approach better preserves the APL metric. The tables presented above demonstrate that the parameters measuring the utility of data after anonymization vary depending on the chosen method or technique, particularly in terms of how the graph is attacked during the anonymization procedure.

This difference is observed in anonymizations performed on the edges rather than focusing on the intricacies of individual nodes, regardless of their types. Consequently, the selection of evaluation parameters is determined based on specific requirements. Hence, it is evident that each article in the comparative analysis employs distinct evaluation parameters.

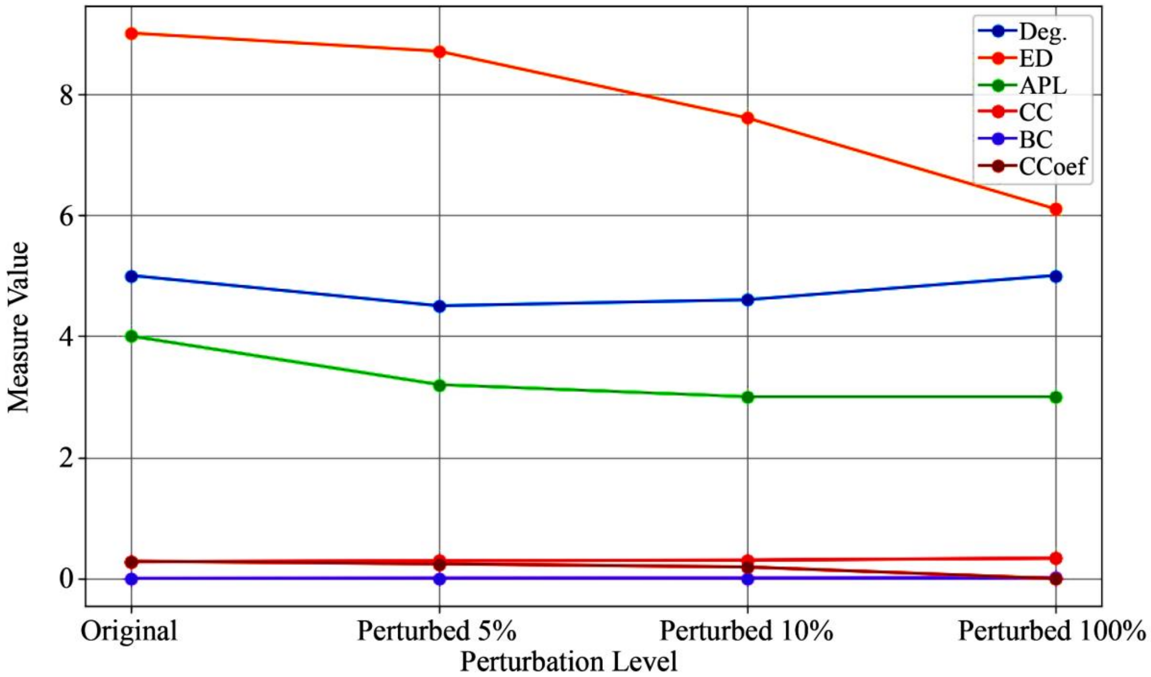


Fig. 4 Effect of anonymization on graph measurements on Enron data (Table 3)

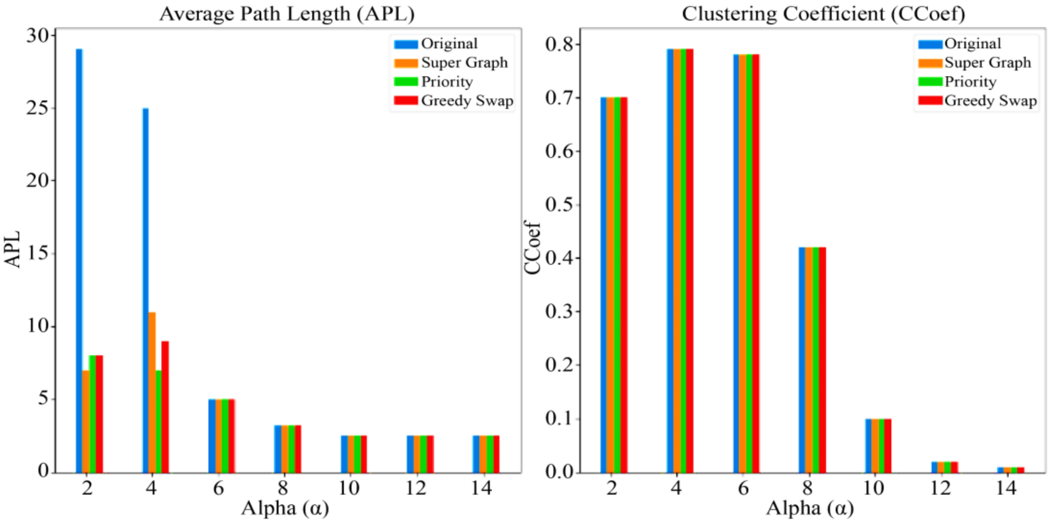


Fig. 5 Average path length and clustering coefficient for the spectrum of greedy Swap graphs (Table 4)

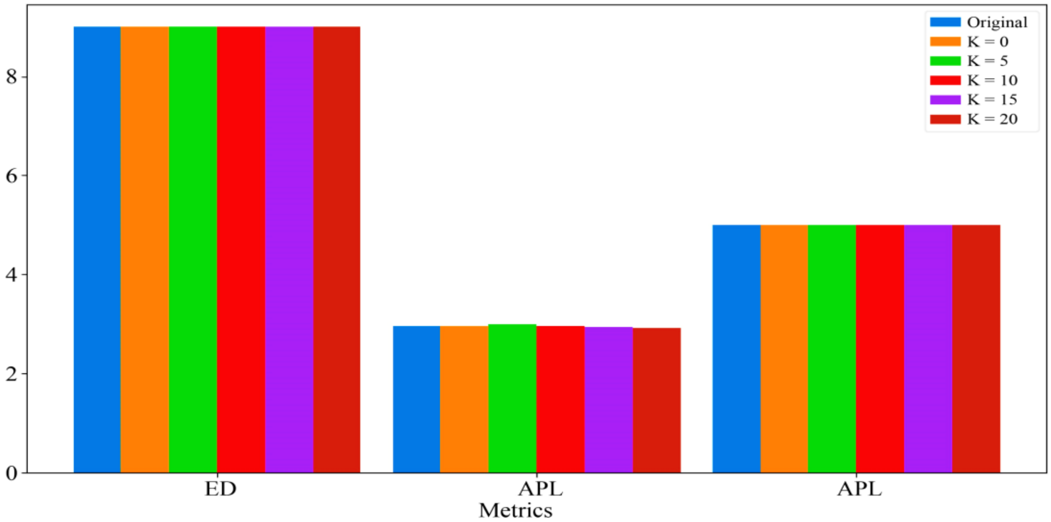


Fig. 6 Average path length, effective diameter, betweenness centrality and radius measures for different k (Table 5)

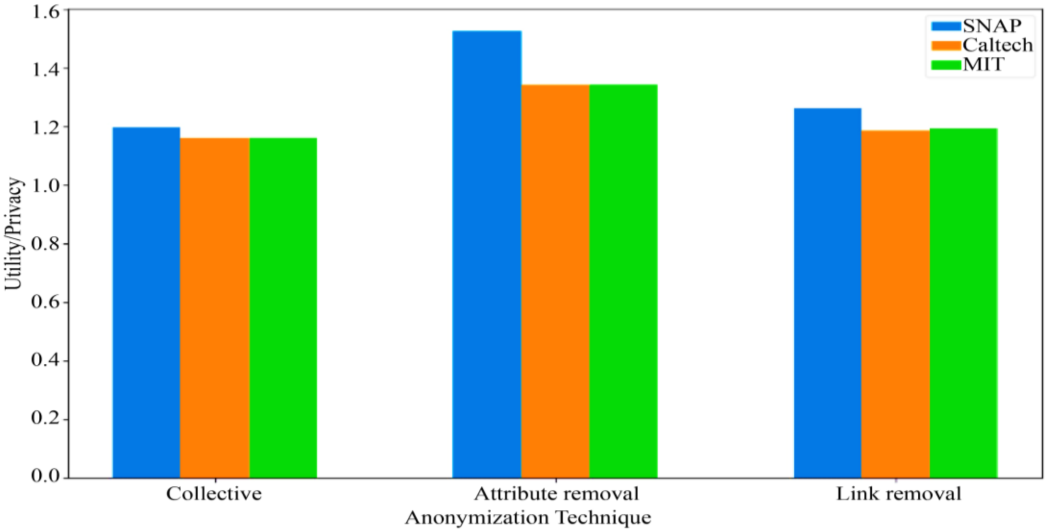


Fig. 7 Data utility and privacy measurement under link removal, attribute removal and collective methods (Table 6)

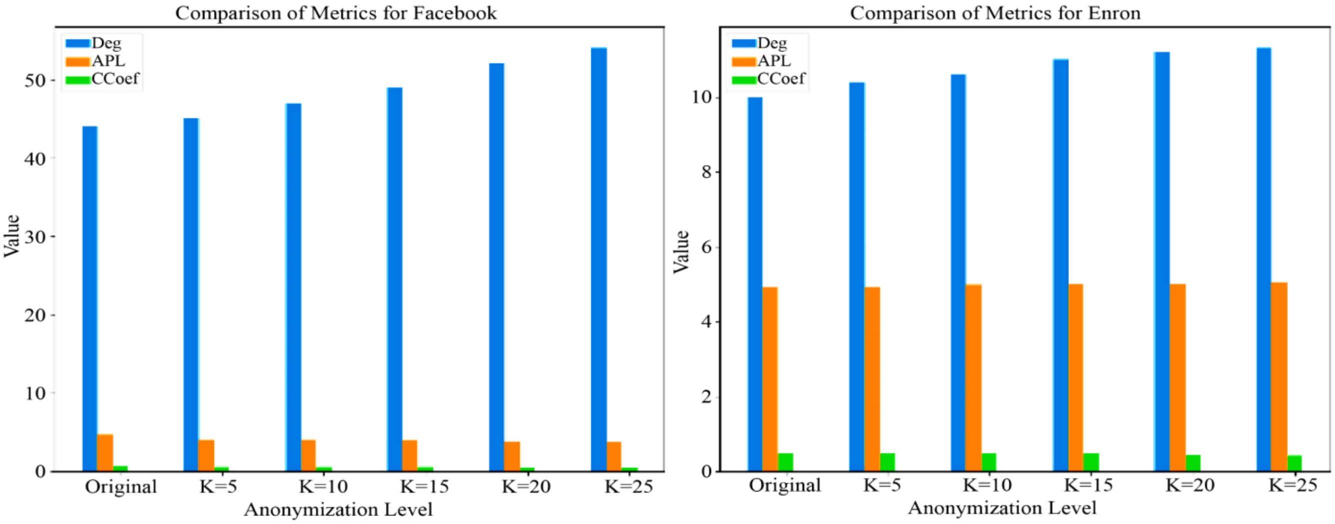


Fig. 8 Utility of the anonymized social network (Table 7)

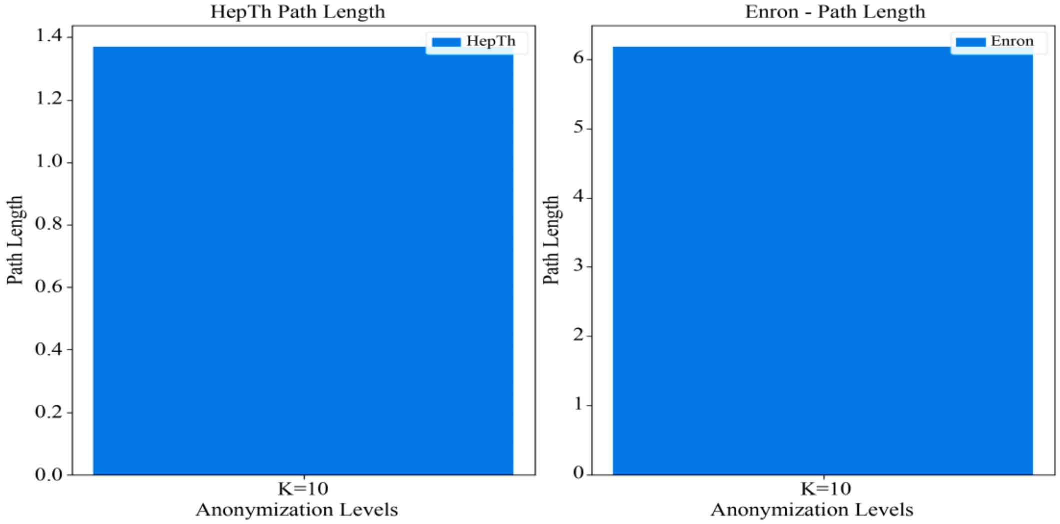


Fig. 9 Shortest-path length distribution Enron and HepTh graphs (Table 8)

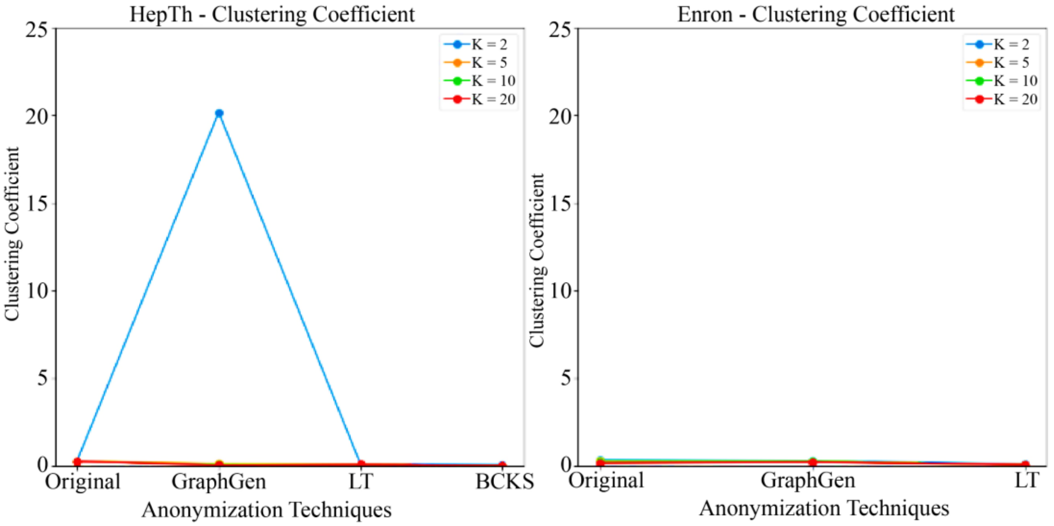


Fig. 10 Clustering coefficient of Enron and HepTh graphs (Table 8)

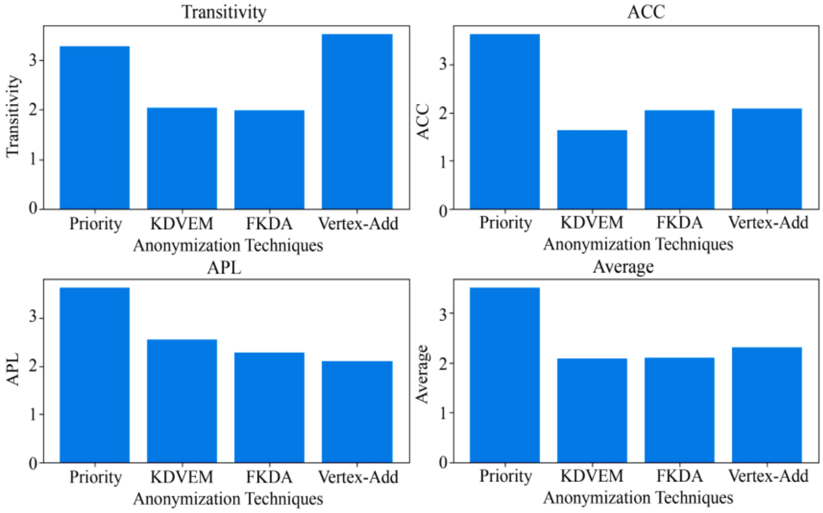


Fig. 11 Relative performance of different algorithms by metric (Table 9)

Table 10. Comparison of datasets used and evaluation metrics

Author	Database	Evaluation
[15]	Enron	• Degree • Effective diameter • Average path length • Closeness centrality • Betweenness centrality • Clustering coefficient
[16]	Enron NetTrace HOT Power-Law Mesh HepTh Tree	• Average path length • Clustering coefficient
[56]	Facebook	• Effective diameter • Average path length • Radius • Betweenness centrality
[48]	SNAP Caltech MIT	• Collective • Attribute Removal • Link Removal
[49]	Facebook Enron	• Degree • Average path length • Clustering coefficient
[14]	HepTh Enron	• Path length • Clustering coefficient
[31]	Ca-HepTh Enron Ca-GrQc Ca-AstroPh Ca-CondMat	• Transitivity • ACC • Average path length • Average

12. Conclusion

In this paper, a comprehensive study on data anonymization in social networks was undertaken. The study began with a detailed examination of social network representations, followed by the introduction of key parameters to assess data utility post-anonymization, which is crucial for evaluating the effectiveness of anonymization methods. Different types of anonymization attacks were also discussed. The various existing anonymization techniques were classified into three categories: graph modification, generalization or clustering, and differential privacy, with distinct methods identified and examined for each category.

A comparative study of highly cited articles was conducted, evaluating the proposed methods based on extracted evaluation parameters. This comparative analysis, as detailed in Section 5, highlighted the diversity in how evaluation parameters are presented across different studies, either in tables or graphs.

The research demonstrated that integrating the strengths of graph modification, generalization, and differential privacy techniques leads to superior outcomes. This comprehensive study addresses gaps and limitations in existing methods by systematically evaluating and combining these techniques. Generalization techniques were found to effectively maintain data utility, graph modification techniques ensured structural integrity, and differential privacy techniques provided robust privacy guarantees with minimal noise introduction.

The study optimized k-degree anonymity using efficient heuristic techniques and innovative edge addition strategies, minimizing structural information loss while maintaining analytical integrity. Additionally, exploring both local and global noise addition strategies allowed for fine-tuning the anonymization process, enhancing privacy protection while preserving data utility for various analytical tasks. Designed for scalability and efficiency, these methods are suitable for large datasets typical in social network analysis. The findings underscore the importance of a balanced and integrated approach to social network data anonymization, paving the way for more secure and useful anonymized datasets in the digital age.

References

[1] Benjamin C.M. Fung et al., *Introduction to Privacy-Preserving Data Publishing*, 1st ed., Chapman and Hall/CRC, pp. 1-376, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[2] Latanya Sweeney, "K-Anonymity: A Model For Protecting Privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557-570, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[3] M. Prakash, and G. Singaravel, "A New Model for Privacy Preserving Sensitive Data Mining," *2012 Third International Conference on Computing, Communication and Networking Technologies*, Coimbatore, India, pp. 1-8, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[4] Mehmet Ercan Nergiz, and Chris Clifton, "δ-Presence without Complete World Knowledge," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 6, pp. 868-883, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ashwin Machanavajjhala et al., "L-Diversity: Privacy Beyond K-Anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yang Xu et al., "A Survey of Privacy Preserving Data Publishing Using Generalization and Suppression," *Applied Mathematics & Information Sciences*, vol. 8, no. 3, pp. 1103-1116, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ruth Brand, *Microdata Protection through Noise Addition, Inference Control in Statistical Databases*, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, vol. 2316, pp. 97-116, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[8] Gregory S. Nelson, "Practical Implications of Sharing Data: A Primer on Data Privacy, Anonymization, and De-Identification," *SAS Global Forum Proceedings*, pp. 1-23, 2015. [Google Scholar]

[9] Stephen E. Fienberg, and Julie McIntyre, "Data Swapping: Variations on a Theme by Dalenius and Reiss," *Privacy in Statistical Databases: CASC Project Final Conference*, Barcelona, Spain, pp. 14-29, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[10] Sean J. Hickman, and Youyi Mao, "Method and System for Data Pattern Matching, Masking and Removal of Sensitive Data," *US20130167192A1*, pp. 1-14, 2013. [Google Scholar] [Publisher Link]

[11] Joana Ferreira Marques, and Jorge Bernardino, "Analysis of Data Anonymization Techniques," *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, vol. 2, pp. 235-241, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] B.K. Tripathy et al., *Privacy and Anonymization in Social Networks, Social Networking*, Intelligent Systems Reference Library, Springer, Cham, vol. 65, pp. 243-270, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[13] Pierangela Samarati, and Latanya Sweeney, "*Protecting Privacy when Disclosing Information: K-Anonymity and its Enforcement through Generalization and Suppression*," 1998. [Google Scholar] [Publisher Link]

[14] Réka Albert, Hawoong Jeong, and Albert-László Barabási, "Error and Attack Tolerance of Complex Networks," *Letters*, vol. 406, no. 6794, pp. 378-382, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[15] Duncan J. Watts, and Steven H. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Letters*, vol. 393, no. 6684, pp. 440-442, 1998. [CrossRef] [Google Scholar] [Publisher Link]

[16] Michael Hay et al., "Resisting Structural Re-Identification in Anonymized Social Networks," *Proceedings of the VLDB Endowment*, vol. 1, no. 1, pp. 102-114, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[17] Michael Hay et al., "*Anonymizing Social Networks*," University of Massachusetts Amherst, pp. 1-17, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[18] Kun J Ray Liu, and Evimaria Terzi, "Towards Identity Anonymization on Graphs," *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, Vancouver Canada, pp. 93-106, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[19] Xiaowei Ying, and Xintao Wu, "Randomizing Social Networks: A Spectrum Preserving Approach," *Proceedings of the 2008 SIAM International Conference on Data Mining*, pp. 739-750, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[20] Lian Liu et al., "*Privacy Preserving in Social Networks Against Sensitive Edge Disclosure*," Technical Report CMIDA-HiPSCCS 006-08, Department of Computer Science, University of Kentucky, 2008. [Google Scholar]

[21] Elena Zheleva, and Lise Getoor, "Preserving the Privacy of Sensitive Relationships in Graph Data," *Privacy, Security, and Trust in KDD, First ACM SIGKDD International Workshop, PinKDD*, San Jose, CA, USA, pp. 153-171, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[22] Alina Campan, and Traian Marius Truta, "A Clustering Approach for Data and Structural Anonymity," *Proceedings of the 2nd ACM SIGKDD International Workshop on Privacy, Security, and Trust in KDD (PinKDD'08), in Conjunction with KDD*, vol. 8, pp. 1-10, 2008. [Google Scholar] [Publisher Link]

[23] Bin Zhou, and Jian Pei, "Preserving Privacy in Social Networks against Neighborhood Attacks," *2008 IEEE 24th International Conference on Data Engineering*, Cancun, Mexico, pp. 506-515, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[24] Bin Zhou, and Jian Pei, "The K-Anonymity and L-Diversity Approaches for Privacy Preservation in Social Networks against Neighborhood Attacks," *Knowledge and Information Systems*, vol. 28, no. 1, pp. 47-77, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[25] Graham Cormode et al., "Anonymizing Bipartite Graph Data Using Safe Groupings," *Proceedings of the VLDB Endowment*, vol. 1, no. 1, pp. 833-844, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[26] Bin Zhou, Jian Pei, and WoShun Luk, "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. [CrossRef] [Google Scholar] [Publisher Link]

[27] Mohammad Soryani, and Behrooz Minaei, "Social Networks Research Aspects: A Vast and Fast Survey Focused on the Issue of Privacy in Social Network Sites," *arXiv*, pp. 363-373, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[28] Kristen Riedt Lefevre, David Johns DeWitt, and Raghu Ramakrishnan, "Incognito: Efficient Full-Domain K-Anonymity," *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, Baltimore Maryland, pp. 49-60, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[29] Qin Liu et al., "Preserving Privacy with Probabilistic Indistinguishability in Weighted Social Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 5, pp. 1417-1429, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[30] Gary William Flake, Robert E. Tarjan, and Kostas Tsioutsiouliklis, "Graph Clustering and Minimum Cut Trees," *Internet Mathematics*, vol. 1, no. 4, pp. 385-408, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[31] Tinghuai Ma et al., "KDVEM: a K-Degree Anonymity with Vertex and Edge Modification Algorithm," *Computing*, vol. 97, no. 12, pp. 1165-1184, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[32] Lars Backstrom, Cynthia Dwork, and Jon Michael Kleinberg, "Wherefore Art Thou r3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography," *Proceedings of the 16th International Conference on World Wide Web*, Banff Alberta Canada, pp. 181-190, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[33] Fionn Murtagh, and Pedro Contreras, "Algorithms for Hierarchical Clustering: An Overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86-97, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[34] Aristidis Likas, Nikos Vlassis, and Jakob J. Verbeek, "The Global K-Means Clustering Algorithm," *Pattern Recognition*, vol. 36, no. 2, pp. 451-461, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[35] Michael Hay et al., "Accurate Estimation of the Degree Distribution of Private Networks," *2009 Ninth IEEE International Conference on Data Mining*, Miami Beach, FL, USA, pp. 169-178, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[36] Shiva Prasad Kasiviswanathan et al., "Analyzing Graphs with Node Differential Privacy," *Theory of Cryptography, 10th Theory of Cryptography Conference*, Tokyo, Japan, pp. 457-476, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[37] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith, "Smooth Sensitivity and Sampling in Private Data Analysis," *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, San Diego, California USA, pp. 75-84, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[38] Francesco Bonchi, Aristides Gionis, and Tamir Tassa, "Identity Obfuscation in Graphs through the Information Theoretic Lens," *Information Sciences*, vol. 275, pp. 232-256, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[39] Alexandre Evfimievski, Johannes Gehrke, and Ramakrishnan Srikant, "Limiting Privacy Breaches in Privacy Preserving Data Mining," *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, San Diego California, pp. 211-222, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[40] Rakesh Agrawal, Ramakrishnan Srikant, and Dilys Thomas, "Privacy Preserving Olap," *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, Baltimore Maryland, pp. 251-262, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[41] Yufei Tao et al., "On Anti-Corruption Privacy Preserving Publication," *2008 IEEE 24th International Conference on Data Engineering*, Cancun, Mexico, pp. 725-734, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[42] Xuesong Lu, Yi Song, and Stéphane Bressan, "Fast Identity Anonymization on Graphs," *Database and Expert Systems Applications: 23rd International Conference*, Vienna, Austria, pp. 281-295, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[43] Ting Yu, and Sushil Jajodia, *Secure Data Management in Decentralized Systems*, Advances in Information Security, Springer New York, NY, vol. 33, pp. 323-353, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[44] Vicenç Torra, Guillermo Navarro-Arribas, and Klara Stokes, *An Overview of the Use of Clustering for Data Privacy*, Unsupervised Learning Algorithms, Springer, Cham, pp. 237-251, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[45] Sean Chester et al., "K-Anonymization of Social Networks by Vertex Addition," *ADBIS 2011, Research Communications, Proceedings II of the 15th East-European Conference on Advances in Databases and Information Systems*, Vienna, Austria, vol. 789, pp. 107-116, 2011. [Google Scholar] [Publisher Link]

[46] Chongjing Sun et al., "Privacy Preserving Social Network Publication Against Mutual Friend Attacks," *2013 IEEE 13th International Conference on Data Mining Workshops*, Dallas, TX, USA, pp. 883-890, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[47] Michael Hay et al., "Resisting Structural Re-Identification in Anonymized Social Networks," *The VLDB Journal*, vol. 19, pp. 797-823, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[48] Faraz Ahmed, Alex X. Liu, and Rong Jin, "Publishing Social Network Graph Eigenspectrum with Privacy Guarantees," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 892-906, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[49] Miguel Ros-Martín, Julián Salas, and Jordi Casas-Roma, "Scalable Non- Deterministic Clustering-Based K-Anonymization for Rich Networks," *International Journal of Information Security*, vol. 18, pp. 219-238, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[50] Xiaowei Ying et al., "Comparisons of Randomization and K-Degree Anonymization Schemes for Privacy Preserving Social Network Publishing," *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, Paris, France, pp. 1-10, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[51] Jordi Casas-Roma, "Privacy-Preserving on Graphs Using Randomization and Edge-Relevance," *11th International Conference Modeling Decisions for Artificial Intelligence*, Tokyo, Japan, pp. 204-216, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[52] Tianchong Gao, and Feng Li, "Sharing Social Networks Using a Novel Differentially Private Graph Model," *2019 16th IEEE Annual Consumer Communications & Networking Conference*, Las Vegas, NV, USA, pp. 1-4, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[53] Peng Liu et al., "Local Differential Privacy for Social Network Publishing," *Neurocomputing*, vol. 391, pp. 273-279, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[54] Tianchong Gao et al., "Local Differential Privately Anonymizing Online Social Networks Under HRG-Based Model," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 4, pp. 1009-1020, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[55] Sepp Hartung, Clemens Hoffmann, and André Nichterlein, "Improved Upper and Lower Bound Heuristics for Degree Anonymization in Social Networks," *13th International Symposium on Experimental Algorithms*, Copenhagen, Denmark, pp. 376-387, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[56] Seyedhashem Hamzehzadeh, and Sayyed Majid Mazinani, "ANNM: A New Method for Adding Noise Nodes which are Used Recently in Anonymization Methods in Social Networks," *Wireless Personal Communications*, vol. 107, no. 4, pp. 1995-2017, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[57] Tianchong Gao, and Feng Li, "Privacy-Preserving Sketching for Online Social Network Data Publication," *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking*, Boston, MA, USA, pp. 1-9, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[58] Ratandeep Kaur, Manisha Sharma, and S. Taruna, "Ensemble Based Model for Privacy in Online Social Network," *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management*, pp. 1277-1282, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[59] Arvind Narayanan, and Vitaly Shmatikov, "Robust De-Anonymization of Large Sparse Datasets," *2008 IEEE Symposium on Security and Privacy*, Oakland, CA, USA, pp. 111-125, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[60] Alina Campan, and Traian Marius Truta, "Data and Structural K- Anonymity in Social Networks," *Privacy, Security, and Trust in KDD: Second ACM SIGKDD International Workshop*, Las Vegas, Nevada, pp. 33-54, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[61] Zhipeng Cai et al., "Collective Data-Sanitization for Preventing Sensitive Information Inference Attacks in Social Networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 577-590, 2016. [CrossRef] [Google Scholar] [Publisher Link]