IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 4, 2020, accepted November 17, 2020, date of publication December 18, 2020,
date of current version January 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3045700

Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey

ABDUL MAJEED AND SUNGCHANG LEE, (Member, IEEE)
School of Information and Electronics Engineering, Korea Aerospace University, Goyang 10540, South Korea
Corresponding author: Sungchang Lee (sclee@kau.ac.kr)

ABSTRACT Anonymization is a practical solution for preserving user's privacy in data publishing. Data owners such as hospitals, banks, social network (SN) service providers, and insurance companies anonymize their user's data before publishing it to protect the privacy of users whereas anonymous data remains useful for legitimate information consumers. Many anonymization models, algorithms, frameworks, and prototypes have been proposed/developed for privacy preserving data publishing (PPDP). These models/algorithms anonymize users' data which is mainly in the form of tables or graphs depending upon the data owners. It is of paramount importance to provide good perspectives of the whole information privacy area involving both tabular and SN data, and recent anonymization researches. In this paper, we presents a comprehensive survey about SN (i.e., graphs) and relational (i.e., tabular) data anonymization techniques used in the PPDP. We systematically categorize the existing anonymization techniques into relational and structural anonymization, and present an up to date thorough review on existing anonymization techniques and metrics used for their evaluation. Our aim is to provide deeper insights about the PPDP problem involving both graphs and tabular data, possible attacks that can be launched on the sanitized published data, different actors involved in the anonymization scenario, and major differences in amount of private information contained in graphs and relational data, respectively. We present various representative anonymization methods that have been proposed to solve privacy problems in application-specific scenarios of the SNs. Furthermore, we highlight the user's re-identification methods used by malevolent adversaries to re-identify people uniquely from the privacy preserved published data. Additionally, we discuss the challenges of anonymizing both graphs and tabular data, and elaborate promising research directions. To the best of our knowledge, this is the first work to systematically cover recent PPDP techniques involving both SN and relational data, and it provides a solid foundation for future studies in the PPDP field.

INDEX TERMS Privacy preserving data publishing, anonymization, privacy, utility, relational data, graphs data, social networks, relational and structural anonymization, information privacy, adversary.

I. INTRODUCTION

Most organizations such as hospitals, banks, insurance companies, and supermarkets collect relevant customers/ subscribers data to improve service quality (SQ). Apart from these physical organizations, an excessive amount of user's data is collected by the virtual platforms such as social networks (SN) service providers due to the extensive use of SN all around the world. With the significant advancement in the information and communication technologies (ICT), SNs enable people to interact with their friends, make new friends, seek information about the relevant subject matter or jobs, spread reliable information at low cost, and also to entertain themselves by watching digital contents. Meanwhile, the SNs collect and store the relevant data about their users during the service provisioning, and at the time of account creation (i.e., joining the SNs). This collected data often contains information about the user's activities, demographics, finance, hobbies, location, perceptions, interests, preferences, political and religious views, online communities, and opinions. Furthermore, most users readily post other valuable data including preferences in music, viewing choices, and social problems such as an epidemic outbreak. Research has shown that analysis of this collected data with advanced data mining tools can assist organizations in improving SQ significantly. For instance, it allows them to understand social trends, people's sentiments and behaviors, and factors causing a certain disease outbreak. Accordingly, such information can be

The associate editor coordinating the review of this manuscript and approving it for publication was Longxiang Gao.

leveraged for many scientific or business objectives including targeted advertisement, relevant content recommendations, and effective decision making [1], [2]. Although the data sharing brings innovation and enables better decision making, it may also jeopardize the privacy of users due to the existence of sensitive information in the data [3].

Before publishing the users' data with the researchers or third parties, data owners ensure that the user's private information privacy is protected. This is typically done via data anonymization, which transforms the original data by applying some operations on it to effectively protect user's privacy without degrading the anonymous data utility [4]. Privacy preserving data publishing (PPDP) provides set of models, tools, and methods to safeguard against the privacy threats that emerge from the data releasing with data miners or analysts [5]. In recent years, PPDP has received considerable attention from the research community, and many approaches have been proposed for both SN and tabular data anonymization [6]–[10]. There are two famous settings of PPDP, non-interactive and interactive [11]. In the former setting, the data owner publishes the complete dataset in an anonymized form after applying some modifications on the original data. However, in the later setting, the data owner does not publish the whole data set in a sanitized form like the former setting. Instead, data owner provides an interface to the data miners through which they may pose different statistical queries about the related data and get (possibly noisy) answers. The k-anonymity model [12], and its ramifications are most widely used in the non-interactive setting of PPDP [13]–[17]. These approaches apply some modifications on the original values of quasi identifiers, and protect the user's privacy by making information less-specific. The differential privacy (DP) [18], and DP based approaches are mostly used in an interactive setting of PPDP [19]–[21]. Meanwhile, some studies have reported the DP based approaches for non-interactive setting [22]. Both k-anonymity and DP based anonymization approaches, and their improved versions have been extensively used in the PPDP.

In recent years, SN data is also published with the data-minders for accomplishing multiple scientific and business objectives due to the phenomenal growth in SN use around the globe [23]. SNs data is mostly in a graph G form, and it provides unprecedented opportunities for advanced data analytics. A social graph data $G(U, V)$, where U is the list of users and V represents the set of edges modelling the relationship between the users. The social connection among users in SNs can be of different types such as friend, sibling, and lover. Generally, each user in a SN has two types of the social connections. Among these connections, there is a set of public connection (V_p), and other connection $V_s = V \setminus V_p$ that are set private by the users, which needs privacy protection. Aside from the V_s privacy protection, there are many aspects in which SN users want privacy protection such as the sensitive attribute (SA), online groups affiliations, and locations. Researchers have extended the concepts used for the tabular data anonymization to protect SN's users privacy [24]–[26]. The two popular anonymization approaches used for the SN data are: naive and structural anonymization. In naive anonymization, only social link structure is published by removing the edges and nodes labels from the G. However, Backstrom *et al.* [27] suggested that naive anonymization is prone to identity disclosure because the structure of the released graph may reveal the identity of the individuals corresponding to the nodes. In contrast, the structural anonymization approaches modify the structure of G to effectively protect the user's privacy. These approaches add new edges, vertices, and/or modify the existing G structure to fulfill the privacy and utility requirements. For instance, in the k-degree anonymity [28], the G containing users and their relationships is modified in such a way that each user U in a G has the degree k. In some cases, the sensitive information (i.e., link information) is removed from the G for some users during SN data anonymization. Zheng *et al.* [29] proposed a framework for preserving sensitive link information privacy in SN data anonymization. Aside from the edges and vertices addition/deletion, advanced techniques such as edges switch and rotation have also been proposed to solve the users' privacy problems in SN data anonymization [30].

The existing surveys related to PPDP cover important aspects such as anonymization techniques, anonymization operations, privacy models, data anonymity frameworks, and evaluating metrics employed by the PPDP mechanisms. Fung *et al.* [31] study systematically summarized and evaluated different approaches used in the PPDP. Rajendran *et al.* [32] explained three prominent and most widely used anonymization models in a medical field, namely k-anonymity, ℓ-diversity, and t-closeness. Gkoulalas *et al.* [33] presented a comprehensive survey about the privacy threats and privacy models used in the PPDP. The authors provided details about fourty-five anonymity algorithms in their study. Tran *et al.* [34] presented a detailed survey about the privacy-preserving big data analytics. The authors explained the related studies and provided details about various PPDP practical scenarios that needs further development from research community. In addition, researchers have presented surveys about the PPDP techniques used in the big data era [35], [36]. A few surveys address the SN data publishing problems, but only considering possible breaches and briefly mentioning the privacy problems that emerge from SN users' data publishing. Yang *et al.* [37] explained about the attack models and countermeasures in SN data publishing. The authors surveyed and categorized the PPDP algorithms into two categories, namely anonymization and DP. Siddula *et al.* [38] provided a survey about the privacy models and methods used for the SN user's privacy protection. The authors explained the mechanism used for the edges and nodes privacy protection in publishing G. Zho *et al.* [39] presented a survey about the anonymization techniques used for SN data, and discussed the challenges involved in the G anonymization compared to the relational data anonymization. Zheleva *et al.* [40]

presented a survey about privacy in SNs and statistical inference techniques. Abawajy *et al.* [41] presented a review of anonymization techniques employed on SN data, and privacy attacks and risks. Furthermore, some surveys related to across SNs user identification [42], [43], edges and vertex modifications techniques [44], and SN application specific scenarios have been reported in the literature [45].

This paper presents a comprehensive survey on recent anonymization techniques used for both SN and relational data publishing. Specifically, our review explains anonymization approaches related to the information privacy protection. The contributions of this review article in the field of PPDP is summarized as: (i) it presents state-of-the-art anonymization techniques used for both SN (i.e., social graphs) and relational (i.e., tabular) data, and fundamental concepts and ideas related to tables and graph data anonymization; (ii) it systematically categorizes the existing anonymization techniques into relational and structural anonymization, and presents an up-to-date thorough review on existing anonymization techniques and metrics used for their evaluation; (iii) it describes the anonymization techniques that have been proposed to solve privacy problems in application-specific scenarios (e.g., collaborative filtering, topic and context modeling, and community clustering etc.) of the SNs; (iv) it presents various methods and items that are exploited by malevolent adversaries for user's re-identification across SNs; (v) it explains various challenges faced by researchers while devising new anonymization methods for tabular and SN data; (vi) it provides new insights on the privacy problems in future computing paradigm that will be helpful in devising more secure anonymization methodologies; and (vii) it discusses promising future research directions in the field of the PPDP that need further development and research from both academia and industry. Through this comprehensive overview, we hope to provide a solid foundation for future studies in the PPDP area.

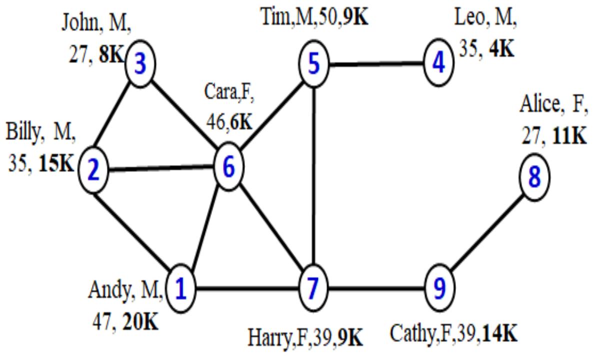
The remainder of this review paper is organized as follows. Section II explains the background regarding privacy types, tabular and graph data overview, types of the user's attributes, operational utility levels of the SN data, privacy areas in the SN, and privacy threats that occur in both SN and tabular data publishing. Section III presents the dilemma of PPDP, and explains its principal concepts and phases. Section IV discusses the state-of-the-art relational anonymization techniques used for tabular data. Section V discusses the recent structural anonymization techniques used for the SN data. The summary of the research contents presented in this article, and discussion about the privacy problems in the future computing paradigm are provided in Section VI. Section VII discusses promising open research directions/problems in the PPDP area. Finally, Section VIII concludes the paper.

II. BACKGROUND

Privacy is all about keeping personal information away from the public access. Privacy is needed for the personal autonomy, individualism, and respect. There are four types of the privacy such as information, bodily, territorial, and communication [46]. Information privacy is about collecting, managing, analyzing, and publishing the personal data. Bodily privacy is related to the physical harms from any kind of invasive procedures/measures. Communication privacy refers to any form of communication such as phone calls or e-mails privacy. Territorial privacy refers to placing boundaries on irruption into a locality. This survey focuses on the information privacy, which encompasses systems/infrastructures that collect, analyse, process, and publish user's data. Concisely, we present the various anonymization approaches that were proposed to anonymize users' data that can be either in the form of graph or table. The overview of the user's data in relational and graphs form is presented in Figure 1. In relational data (Figure 1 (a)), each tuple contains four types of attributes about users, direct identifiers (DI), non-sensitive attributes (NSA), quasi identifiers (QIs), and sensitive attribute (SA). In contrast, the SN data shown in Figure 1 (b) represents the users information via nodes/edges labels. The relevant background about both types of data (i.e., tabular and graphs) is covered in subsequent subsections.

	DI	NSA	Quasi Identifiers (QIs)				SA
ID	Name	Height	Age	Zip Code	Martial Status		Crime
1	Joe	5	29	32042	Separated		Murder
2	Jill	4	20	32021	Single		Theft
3	Sue	6	24	32024	Widowed		Traffic
4	Abe	5	28	32046	Separated		Assault
5	Bob	7	25	32045	Widowed		Piracy
6	Amy	6	23	32027	Single		Indecency

(a) Microdata table of the criminal records(Tabular data).

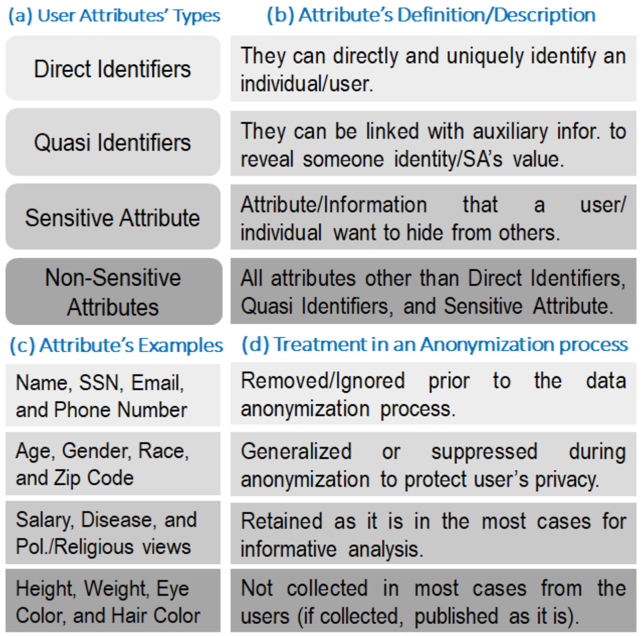


(b) Friends network of a consulting firm (Graph data).

FIGURE 1. Overview of the relational (i.e., tabular) and social network (i.e., graphs) users data.

A. BACKGROUND ABOUT THE RELATIONAL/TABULAR DATA

The original user's data D is considered as a private table which consists of multiple records (i.e., tuples). Each record/tuple contains four types of attributes, and every record has a unique id as shown in Figure 1(a). The detailed overview of user's attributes along with examples and their treatment in an anonymization process is illustrated in Figure 2. Based on the types of user's attributes listed in Figure 2, there exists three classes of privacy threats that can occur during published data analysis [47], which are explained below:



(a) User Attributes' Types	(b) Attribute's Definition/Description
Direct Identifiers	They can directly and uniquely identify an individual/user.
Quasi Identifiers	They can be linked with auxiliary infor. to reveal someone identity/SA's value.
Sensitive Attribute	Attribute/Information that a user/ individual want to hide from others.
Non-Sensitive Attributes	All attributes other than Direct Identifiers, Quasi Identifiers, and Sensitive Attribute.

(c) Attribute's Examples	(d) Treatment in an Anonymization process
Name, SSN, Email, and Phone Number	Removed/Ignored prior to the data anonymization process.
Age, Gender, Race, and Zip Code	Generalized or suppressed during anonymization to protect user's privacy.
Salary, Disease, and Pol./Religious views	Retained as it is in the most cases for informative analysis.
Height, Weight, Eye Color, and Hair Color	Not collected in most cases from the users (if collected, published as it is).

FIGURE 2. Description about the types of user's attributes and their handling in an anonymization process.

- *Identity disclosure (i.e., unique identification)*: It is a well-known privacy threat in the PPDP. It occurs when an adversary can correctly associate an individual in a privacy preserved published dataset. Generally, an attacker use the information gathered from external sources (i.e., voter registration list, online repositories, and factual information) to identify an individual uniquely.
- *Attribute disclosure (i.e., private information disclosure)*: This type of privacy threat occurs when an individual is linked with the information about his/her SA. For example, the information can be the person's value for the SA (i.e., crime in Figure 1(a), or salary in Figure 1(b)). This type of threat can be easily launched in imbalanced datasets (i.e., the datasets lacking heterogeneity in SA's values.).
- *Membership disclosure (i.e., presence/absence disclosure)*: This threat occurs when an adversary can deduce that an individual's record is present/absent in the published dataset with a very high probability. Researchers have reported many interesting scenarios in which the protection from the membership disclosure is imperative [48], [49].

To protect the privacy of users in the published dataset, data owners can apply one of the following anonymization operations on the original user's data given in a tabular form [50].

- *Generalization*: This operation transforms the original QI's values into less-specific but semantically-consistent values during anonymization process. For example, the value 25 of a QI age can be generalized with an interval $[25 - 30]$ or < 30. This operation relies on the taxonomy of each QI. The existing generalization schemes can be classified to five types such as sub-tree generalization, full domain generalization, unrestricted sub-tree generalization, cell generalization, and multidimensional generalization.
- *Suppression*: This operation hides an original value of a QI with a special value (i.e., '*'). For example, to anonymize the value 25 of a QI age using suppression operation, 5 can be replaced with an asterisk, resulting $2*$ as the suppressed value of QI. Record, value, and cell suppression are the three most widely used suppression variants in the tabular data anonymization.
- *Permutation*: In this operation, the records are partitioned into several groups, and values of the SA are shuffled within each group. Hence, the SA and QIs relationships are de-associated within each group. This operation may yield inaccurate analysis in terms of anonymous data utility, but user's privacy is significantly preserved.
- *Perturbation*: In this operation, the original data values are replaced with some synthetically generated values. Moreover, the synthetic values are generated in a way that statistical information do not differ much in both datasets (i.e., real and synthetically generated datasets).
- *Anatomization*: This operation does not apply any modifications on the original data values and instead QIs and SA are separated into two tables. By doing so, the association between QIs and SA is broken, and data is released as QIs and SA tables separately. In some cases, the SA table contains the SA's values and their frequency in the anonymized dataset for privacy preservation effectively.

Furthermore, in some cases, more than one operation are jointly used to anonymize users data set. Data publication can be done either one or multiple times depending upon the requirements and information consumer's needs. Typical data publication scenarios include one and multiple time publication of the micro data. In the former scenario, the data is published once, and no re-publication is made even after the changes in the original data. In contrast, in the multiple releases, the anonymous data is re-published even after a single operation (insert, update, delete) that changes the original data. Both scenarios have their own advantages and disadvantages for data owners and information consumers, respectively. After DIs and NSA removal from D as a standard PPDP practice, the D contains only two types of user attributes, QIs and SA, represented as $D\{Q, S\}$. We can use set $Q = \{q_1, q_2, \ldots, q_p\}$ to denote the QIs, where q_p is one type of QI such as gender or zip code. Set S represents the SA which can be of single type (i.e., disease) or multiple types (i.e., disease and salary) depending upon the scenario. The multiple SA scenario is getting significant attention from the research community in recent years [51]. Plenty of solutions have been proposed for the relational data anonymization considering the available data, SA scenarios (single, multiple), and the PPDP settings. We explain most famous anonymization solutions and their variants proposed for the tabular data anonymization in Section IV.

B. BACKGROUND ABOUT THE SOCIAL NETWORKS/GRAPHS DATA

The SN user's data can be modeled with a graph G, represented as $G(V, E, A)$. It consists of user set V, social connections (i.e., friendship links) set E, where $E \subseteq V \times V$, and the set A of users' attributes. Set A, where $A = \{Q, S\}$ contains QIs and SA of the SN users, respectively. The overview of G along with the relevant details is depicted in Figure 1(b). In literature, V, E, and A are also referred as nodes, edges, and labels (i.e., user profile), respectively. The SN data is extremely useful for many analytical purposes, the operational utility of the SN data can be classified into three levels (l_1, l_2, l_3), as outlined earlier [52].

- l_1*: Exposure of the graph structure only*: In this exposure level, the data owners (e.g., SN service providers) only publish the graph structure (i.e., all profiles/labels information is removed prior to data publication). Thus, the data-miners/analysts can analyze only the graph structure without any concrete information about the user's profiles.
- l_2*: Exposure of the nodes' profiles*: In this exposure level, the data owner publishes the profiles of nodes/users but hides the graph structure. For instance, the node's data is stored in a table/matrix, and released for the analytics and data mining purposes.
- l_3*: Exposure of both graph structure and profiles of nodes (e.g., users)*: In this exposure level, the data owner exposes both the graph structure and the nodes' profiles after applying some modifications to the G's structure and users' profiles. This level offers much higher utility to the legitimate information consumers compared to the previous two levels.

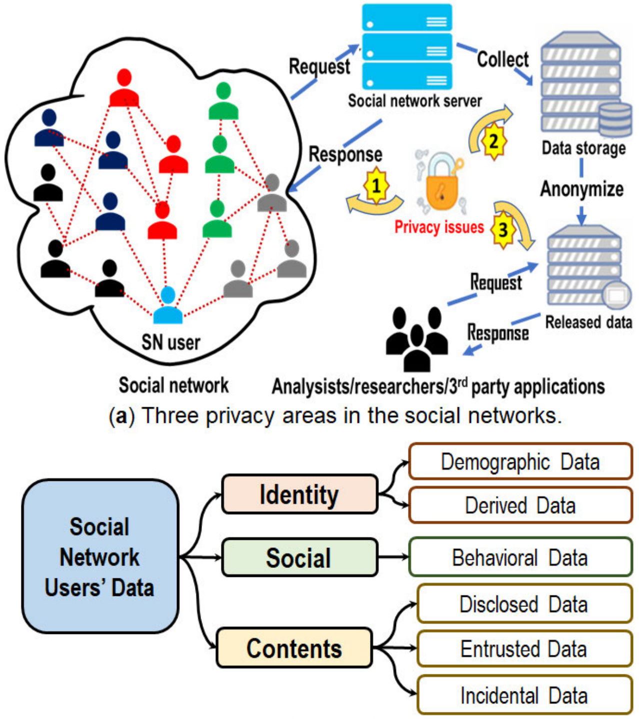
Although SN data publishing is invaluable for accomplishing multiple research and business objectives, the SN data publishing can confront with the privacy threats of several types. Four well-known privacy threats that can happen after G' publishing are summarized below.

- *Identity disclosure (i.e., node re-identification)*: It occurs when an adversary can accurately associate/identify an individual from a privacy preserved published graph. For example, Tim (5*th* node) identification by an adversary from the anonymous version (i.e., removing all nodes labels.) of a G given in Figure 1(b) is an example of identity disclosure (i.e., node re-identification).
- *Edge disclosure (i.e., relationship/connection disclosure)*: It reveals the relationship between users. For example, a patient-doctor relationship can be highly sensitive, and it must be protected. If a doctor is known to be an expert in cancer treatment, the relationship disclosure can occur with inference that patient might be infected with a cancer.
- *Content disclosure (i.e., vertex/edge labels disclosure)*: It occurs when a sensitive label associated with an edge or vertex is revealed from a G', and this reveled label can be directly associated with a specific individual in an original graph (i.e., G).
- *Affiliation link disclosure (i.e., whether a person* $v \in / \notin$ *to a particular affiliation group* h*)*: It occurs when a link between a user v and an affiliation group h is revealed with confidence $\geq t$, and this revealed link can be directly associated with a v. The h can be prosecuted political group or a group centered around an unconventional user's preference (i.e., sexual/drugs preferences).

To protect the privacy of users in SN data publishing, data owners can apply one of the following anonymization operations on the G prior to its publication with the analysts [53].

- *Graph modification*: This operation changes the original graph's structure by adding/deleting edges or vertices. The criteria about the graph's elements addition/deletion depends on the objectives of data publishing, and privacy protection level data owners want. Typical graph modifications techniques are of two types-constrained and non-constrained graph modifications.
- *Graph generalization*: This operation does not modify the graph structure, instead, it cluster the nodes and edges into super nodes and edges. This operation works in iterations, and in each iteration the connections between the super nodes and edges are established.
- *Graph computation (i.e., privacy-aware computation of graphs)*: This operation computes the output data in response to the data miner queries. The G is not released with the data miners, instead, the graph properties (degree, centralities, size, and other useful information) are computed and released.
- *Hybrid operation/approach*: This operation employs combination of two or more anonymization operations simultaneously. For example, the graph generalization and modifications can be jointly used to anonymize G to produce G'.

Privacy in SN data publishing has become an active area of research in recent years. Pham et. al [54] explained that SN users privacy can be breached through three different ways such as user's activities on the SNs sites, stored users data in the SNs service providers' servers/databases, and privacy preserved SN published data. All three privacy areas in SN are visually presented in Figure 3 (a). The privacy in the first two areas can be preserved through guidelines, encryption and watermarking techniques. Meanwhile, the privacy in the third area can be guaranteed only by devising/implementing new anonymization mechanisms. This article covers recent concepts, methods, and solutions concerning the third area of privacy in the SNs (i.e., privacy preserved graph publishing).



(a) Three privacy areas in the social networks.

(b) Types of users data collected/processed in SNs.

FIGURE 3. Overview of three privacy areas (adopted from [54]), and types of users data collected and processed in SNs.

The user data available on the SN sites is comprised of three types of information [55]. The taxonomy of the users data present on the SNs is depicted in Figure 3(b). The first type of data is related to user identity including the demographics and derived data (i.e., age). The second type of information is related to user's socialism (i.e., friends, friends-of-friends, activities, and online communities joined by the users etc.). The third type is related to the contents users create on the SNs sites. It has three types: disclosed, entrusted, and incidental data. Disclosed data is readily made available by the SNs users. Entrusted data includes the data posted by a user on another user's profile (i.e., comments). The last type of data is the information collected/posted by other users on someone's profile/wall. All these data sources are invaluable for detailed analytics and appropriate information collection/analysis.

Aside from the fact that all necessary and private information needs protection from the malevolent adversaries, some de-anonymization attacks can be launched on the SNs published data to infer someone's real identity/private-information. The revelation of such information can result in unknown user profiling, thus targeting unsuspecting users with far-reaching implications including targeted marketing, obtaining travel visas based on race or political viewpoints, deportation, or identity theft. Therefore, mining and sharing SNs data must not invade user's privacy. To safeguard user's privacy in SN data publishing, many privacy preserving graph publishing (PPGP) methods have been proposed. We describe most recent anonymization solutions proposed for the PPGP in Section V.

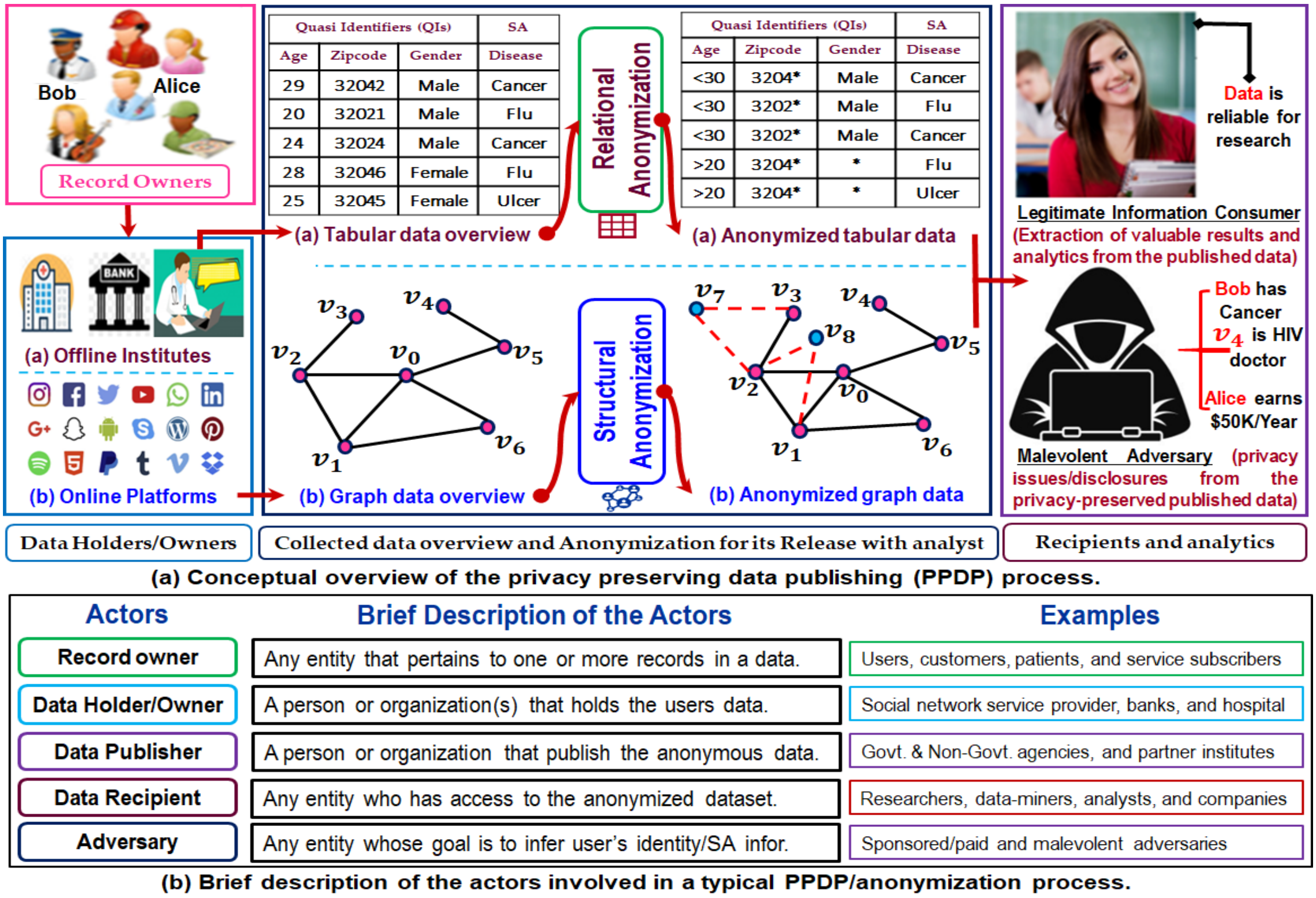
III. OVERVIEW OF PRIVACY PRESERVING DATA PUBLISHING AND ITS FUNDAMENTAL PHASES

Privacy preserving data publishing (PPDP) provides set of tools, methods, solutions, and frameworks for sharing valuable information with analysts/researchers without jeopardizing user's privacy. PPDP has been extensively studied in the literature for protecting different aspects of user's private information during published data analytics. Data is shared with the analysts/ researchers for extracting the embedded knowledge from it. Meanwhile, the anonymous data sharing after applying anonymization operations, as outlined earlier (see subsections II-A, II-B), adversaries can still infer the information about user's identity or private information by leveraging the auxiliary information gathered from external sources. Therefore, many studies have suggested the users' privacy-protection in all stages of the information processing cycle (i.e., collection, storage, processing, release, and destruction/archival) [56]. The conceptual overview of the PPDP process is presented in Figure 4 (a). In this conceptual overview, we mainly present the overview of data collected from the individuals, different actors involved in the anonymization scenario, anonymization techniques applied on respective data, anonymous data to be published for analytics/mining purposes, and privacy breaches that can occur during published data analytics. The typical PPDP scenario involves five types of actors [57]. The brief description about each actor along with examples is summarized in Figure 4 (b). In some cases, one actor can perform multiple roles in the PPDP scenario. For instance, data holders/owners hold the collected data, and perform anonymization for its releasing with analysts. Sometimes due to the lack of computing resources or knowledge the data holder outsources the collected data for anonymization/publishing. Hence, the roles of data holder and data publisher can correspond to two distinct/same actors.

The typical PPDP process encompassed of the five fundamental phases: (i) data collection from the individuals; (ii) collected data storage, understanding, and preparation for the anonymization; (iii) user's data anonymization; (iv) anonymous data releasing/ publishing; and (v) published data analysis for extracting embedded knowledge. Brief details of each phase of the PPDP is presented below.

A. PHASE 1: DATA COLLECTION FROM THE INDIVIDUALS

In the first phase of PPDP, relevant data from the individuals/users is collected. Due to significant technological development in recent years, the amount of data generated by sensor networks, SNs sites, healthcare applications, internet, online banking, SN integrated third party applications, and many other offline/online companies is drastically increasing. This data is collected from individuals directly or via smart devices (i.e., cell phones, laptops, and notepads etc.). For instance, if a patient visits a hospital for diagnosis, his/her


Bob
Alice
Record Owners
(a) Offline Institutes
BANK
(b) Online Platforms
Data Holders/Owners
Relational Anonymization
Structural Anonymization
v_3
v_4
v_2
v_0
v_5
v_1
v_6
v_7
v_8
(b) Graph data overview
(b) Anonymized graph data
Collected data overview and Anonymization for its Release with analyst
Data is reliable for research
Legitimate Information Consumer (Extraction of valuable results and analytics from the published data)
Bob has Cancer
v_4 is HIV doctor
Alice earns $50K/Year
Malevolent Adversary (privacy issues/disclosures from the privacy-preserved published data)
Recipients and analytics

(a) Tabular data overview

Quasi Identifiers (QIs)			SA
Age	Zipcode	Gender	Disease
29	32042	Male	Cancer
20	32021	Male	Flu
24	32024	Male	Cancer
28	32046	Female	Flu
25	32045	Female	Ulcer

(a) Anonymized tabular data

Quasi Identifiers (QIs)			SA
Age	Zipcode	Gender	Disease
<30	3204*	Male	Cancer
<30	3202*	Male	Flu
<30	3202*	Male	Cancer
>20	3204*	*	Flu
>20	3204*	*	Ulcer

(a) Conceptual overview of the privacy preserving data publishing (PPDP) process.

Actors	Brief Description of the Actors	Examples
Record owner	Any entity that pertains to one or more records in a data.	Users, customers, patients, and service subscribers
Data Holder/Owner	A person or organization(s) that holds the users data.	Social network service provider, banks, and hospital
Data Publisher	A person or organization that publish the anonymous data.	Govt. & Non-Govt. agencies, and partner institutes
Data Recipient	Any entity who has access to the anonymized dataset.	Researchers, data-miners, analysts, and companies
Adversary	Any entity whose goal is to infer user's identity/SA infor.	Sponsored/paid and malevolent adversaries

(b) Brief description of the actors involved in a typical PPDP/anonymization process.

FIGURE 4. Privacy preserving data publishing (PPDP): (a) conceptual overview and (b) description of the actors involved in the PPDP scenario.

personal information is collected for the better treatment. Later, the patient's personal information combined with the disease information is saved in the hospital database for the secondary use [58]. Similarly, the account opening process in a bank is subject to the collection of basic as well as sensitive information about the customers. Aside from visiting an organization (i.e., hospital or bank), in some cases, the service providers collect relevant data from their users via questionnaires and interviews.

Recently, due to the significant development in ICTs, majority of service providers have launched their own websites for the data collection from their respective users/customers. Furthermore, in an account creation process on the SNs sites, the service providers collect and store the basic information about each user. In addition, many SNs sites collect the valuable data about users without their consent. For example, they can collect the device information, service consumption temporal information, location information, and many other useful information. Recently, with the inventions of advanced tools and technologies, many infrastructures collect and process graphs data. The nine well-known graph based data collection sources are: (i) social networks, (ii) communication data, (iii) mobility traces, (iv) healthcare and epidemiological data, (v) citation networks, (vi) collaboration network, (vii) web graphs, (viii) autonomous system graphs, and (ix) computer networks. SN data is mostly represented as a G along with the entities (i.e., users) information for modeling/analysis purposes.

B. PHASE 2: COLLECTED DATA STORAGE, UNDERSTANDING AND PREPARATION FOR THE ANONYMIZATION

Once the relevant users' data is collected, the next phase is collected data storage, understanding, and preparation for further operations (i.e., anonymization). Companies are using large scale databases for storing the collected users' data/information. The well-known SN, Facebook uses MySQL database to store/manage many petabytes of data about user's social activities such as shares, comments, and likes. Data understanding involves the analysis of the data types, different data representations, and (*key*, *value*)-pair understanding. Generally, the collected data about users can contain incorrect values, outliers, missing values for some attributes, and incomplete records. Therefore, it needs preparation before the data anonymization process. The data preparation includes: removal of the outliers present in the data which are not appropriate for the analysis and can yield inaccurate results/analysis. Furthermore, it eliminates those records with unknown (i.e., missing) values from the user's data. In some cases, the data processing model/algorithm needs user's information in a specified format. Therefore, the appropriate formatting of the collected data, and enrichment if required for the subsequent steps is performed in the data preparation phase. With the help of data preparation/preprocessing, cleaned data can be obtained which contains complete information about each user, and it can be directly fed into the anonymization algorithm for anonymization.

C. PHASE 3: DATA ANONYMIZATION

Data anonymization is a practical and most widely used solution for protecting user's privacy in data publishing. In tabular data, data anonymization sanitizes the QI's original values to make information less specific for privacy protection and utility enhancements. In contrast, if users' data is given in a G form, anonymization changes the graph structure to protect privacy of users and their associated SA without significantly decreasing anonymous graph's utility. In addition, anonymization can be tailored with the data owner's privacy requirements, legitimate information consumer's utility needs, and objectives of the data publishing. The typical anonymization process includes following four main steps. All four steps are complementary, and can be employed to produce the anonymous table T' from an original table T or anonymous graphs G' from an original graph G.

1) REMOVAL OF DIRECTLY IDENTIFIABLE INFORMATION FROM THE ORIGINAL DATA

At the beginning of the anonymization process, any information that can identify someone directly/uniquely is removed from the data. For example, the name, social security number (SSN), email address, and cell phone number can be linked with someone's real identity. Hence, such information is removed from the data before its anonymization. It can be removed at any stage, but the earlier removal can assist in saving computing power significantly.

2) CHOICE OF THE ANONYMIZATION TECHNIQUE

Many anonymization techniques have been proposed and implemented for different scenarios. In this work, our focus is on the relational (i.e., tabular) data and SN (i.e., graphs) data anonymization. Therefore, we categorize the existing anonymization techniques into two categories: structural and relational anonymization. The former technique is applied to the graph data. In contrast, the latter technique is used for tabular data anonymization. Although some researchers have used the relational anonymization techniques for SN data anonymization. But, due to the significant differences in terms of the information contained in the graphs, relational techniques cannot be directly applied to the SN data. Therefore, the decision about the anonymization technique depends upon the original data representation (i.e., graphs or tables), and objectives of the data publishing.

3) SELECTION OF THE ANONYMIZATION OPERATION

Once the appropriate anonymization technique is chosen, the next step is to employ the relevant anonymization operation to distort the original data values or graph structure. The anonymization operation for each technique are different, as outlined earlier (Subsections II-A, II-B). The selection about the anonymization operation is dependent on the data type, anonymization technique, and privacy and utility objectives. For example, in relational anonymization, the generalization operation retains more semantics of the original data compared to the suppression operation. In contrast, suppression operation is highly appropriate for the user's privacy protection compared to the generalization. In addition, due to the complex structure of a G, changes in the small portion of a G can drastically decrease utility of the G'. Therefore, the selection of the appropriate anonymization operation is made to effectively resolve the privacy and utility trade-off for real-world applications.

4) ENFORCEMENT OF THE CONSTRAINS (IF ANY) IN AN ANONYMIZATION OPERATION

Aside from the selection of appropriate anonymization technique and anonymization operation, some constrains can be enforced by the data owners during data anonymization. Such constrains can be about the privacy and utility thresholds, number of users in an equivalence class, distribution of the user's private values in a class/cluster, and/or the number of connections between users (i.e., degree) in a G'. Such constraints are enforced during data anonymization process. These constraints can be suggested by the data owners or can be derived from the data statistics. In addition, the constraints can be employed by considering the adversaries capabilities and nature of the sensitive information contained in a T or G.

D. ANONYMOUS DATA RELEASING/PUBLISHING

The output of the anonymization process is the anonymous G' or T' depending upon the representation/style of original user's data. This anonymous G' or T' is set to be made publicly available for the analytics. The recipients of the anonymous data can be analysts, researchers, data-miners, analytics firms, and third party applications. Before making the anonymous data publicly available, the data owners perform several checks to verify the user's privacy protection and anonymous data utility levels, respectively. After the detailed checks, the decision about the data release is made by the data owners. In some cases, the data owners do not publish full anonymous data, instead, some parts of the anonymous data are published first. Later, the complete anonymous table or graph is published. In addition, the anonymous data can be shared only with the relevant institutes via emails or posts. However, the anonymous data is generally published over the Internet so that a large number of people can access the published data for the multi-facet analytics. Furthermore, the medium of data sharing and amount of data vary with the setting (i.e., interactive and non-interactive) of the PPDP.

After releasing the data, the data owner has no control over the published data use and distributions. Meanwhile, it is assumed that published data will be used only for the intended purposes, and any problem will be explicitly reported to the data owners. In some cases, the data owners and data publishers can be two different parties. Data publishers release the anonymous data via their own mediums under the publishing agreements with the actual data owners/holders.

E. PHASE 5: PUBLISHED DATA ANALYSIS FOR EXTRACTING EMBEDDED KNOWLEDGE

Once anonymous data has been published, the intended recipients collect it for the analysis. In case of the hospitals data, medical students collect the data for their research. They can perform several kinds of test such as factors causing certain disease, common disease in a certain age-groups people, and symptoms of a particular disease. In addition, the banks data containing information about the loan return rating is valuable for the insurance companies. The SN data is suitable for the digital service providers and marketing firms. Recently, many companies are mining the SN data at large scale for fulfilling their scientific and business objectives. In addition, many third party applications are buying SN user's data for fulfilling their intended objectives. The published data is not only for understanding the causes of some problems, but it can help in devising new rules and patterns that can be useful for marketing purposes. With the help of data mining algorithms, one can analyze the published data for actionable insights. Furthermore, users clustering and analysis is beneficial for recommendations, preferences mining, target marketing, information diffusion, information control, and information trustworthiness. Hence, data sharing has become a routine matter for some companies/organization due to the significant advantages in terms of improved decision making, policy enhancements, trends analysis, forecasting, predictions, and innovation.

Unfortunately, data publishing can jeopardize user's privacy as adversaries can get copy of published data with aim to re-identify people uniquely by leveraging the large amount of auxiliary information obtained from external sources. These information can be gathered from many sources including users' profiles from the SN sites, voter registration list, and mobility traces etc. Generally, the adversaries possess strong programming skills and tools knowledge. Therefore, with the help of auxiliary information, tools understanding, advanced programming skills, and understanding of anonymization methods enable adversaries to breach user's privacy. Due to the advancements in ICTs, the scale and scope of the data breaches is expanding. In recent years, the adversaries are focusing for the users groups information theft for fulfillment of the intended objectives. The group identity theft can lead to the negative perceptions about certain ethnic group, discrimination based on the race/religion, and loan declining to group of people whose previous loan return rating was not satisfactory. Aside from the negative consequences of privacy breaches on the people's life, data owners also lose their users' trust in such circumstance. Therefore, users expect that data owners protect their sensitive information's privacy during data release with the data-miners. Hence, the academicians and researcher are suggesting/devising new anonymization mechanisms to deal with this uprising social problem (i.e., user's privacy protection without significantly impacting data utility/quality).

IV. RELATIONAL ANONYMIZATION TECHNIQUES USED FOR TABULAR DATA ANONYMIZATION

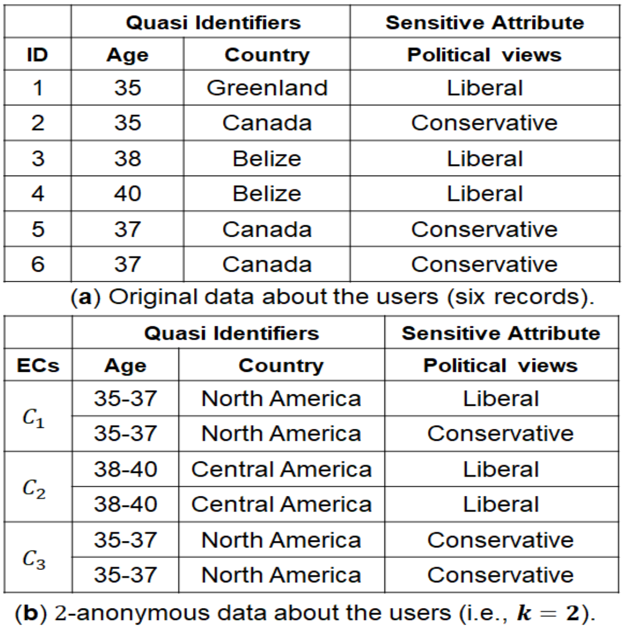
The general concept of relational anonymization is to produce anonymous table T' from an original Table T. Given a T containing p QIs, single/multiple SA (s), and N users, the relational anonymization employs a privacy model/algorithm to modify the original values in such a way that user's privacy can be protected while retaining significant utility in anonymous data for the analysis. The input to the relational anonymization is a tabular data and output is also a tabular data with modified values of user's QIs, and shuffled/equally-distributed values of the SAs. A series of privacy models and algorithms have been proposed so for to anonymize tabular data containing user's basic (i.e., QIs) and private information (e.g., SA). Four well-known and extensively used privacy models for relational data anonymization are summarized below.

A. k-ANONYMITY PRIVACY MODEL

The k-anonymity [12] privacy model is a well-known syntatic privacy model, and it has been extensively used in the tabular data anonymization. Due to the conceptual simplicity, this privacy model has attracted significant attention from the research community, and many variants of this model have been proposed by the researchers for data anonymity. The k-anonymity model [12] protects user's privacy by placing at least k users in an equivalence class (EC) with same QI's values. Hence, the probability of re-identifying someone from T' becomes $1/k$. A table T' satisfies k-anonymity if for every tuple/record t of T' there exist at least $(k-1)$ other tuples with the same QIs in an EC. The value of k is chosen by the data owners depending upon table size and privacy protection level. The k-anonymity privacy model overview is shown in Figure 5. Primarily, the k-anonymity privacy model was devised for the protection of identity disclosure. Meanwhile, the k-anonymity privacy model is insufficient to protect the sensitive information disclosure as shown in ECs, C_2 and C_3 in Figure 5(b). The SA's disclosure in these two ECs based on auxiliary information is 100%. Therefore, an advanced privacy model, named ℓ-diversity was proposed to solve the shortcomings of the k-anonymity privacy model. The utility of the anonymous table T' produced by the k-anonymity model is relatively higher.

B. ℓ-DIVERSITY PRIVACY MODEL

The ℓ-diversity privacy model [59] was proposed to solve the k-anonymity model's limitations. According to this model, an EC satisfies ℓ-diversity property if there are at least ℓ "well-represented" values for the SA. A table T' is said to



	Quasi Identifiers		Sensitive Attribute
ID	Age	Country	Political views
1	35	Greenland	Liberal
2	35	Canada	Conservative
3	38	Belize	Liberal
4	40	Belize	Liberal
5	37	Canada	Conservative
6	37	Canada	Conservative

(**a**) Original data about the users (six records).

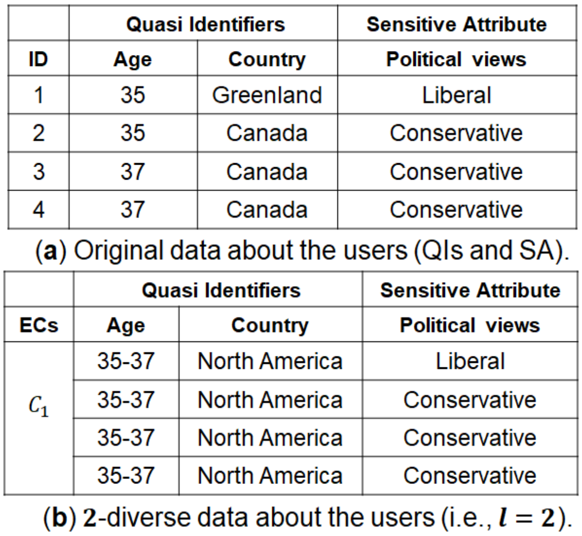
	Quasi Identifiers		Sensitive Attribute
ECs	Age	Country	Political views
C_1	35-37	North America	Liberal
	35-37	North America	Conservative
C_2	38-40	Central America	Liberal
	38-40	Central America	Liberal
C_3	35-37	North America	Conservative
	35-37	North America	Conservative

(**b**) 2-anonymous data about the users (i.e., $\boldsymbol{k} = \mathbf{2}$).

FIGURE 5. 2-anonymity applied to the user's relational data with six records.

have ℓ-diversity, if every EC of the T' is ℓ-diverse. The ℓ-diversity privacy model overview is shown in Figure 6.

The table in Figure 6(b) is an example of 2-diverse partition of the table shown in Figure 6(a). Although, ℓ-diversity privacy model provides superior privacy protection compared to the k-anonymity model by considering the diversity in SA's values, but it does not consider the distribution of the SA's values. Hence, it is prone to the privacy breaches in ECs in which one particular SA value is dominate over others (i.e., C_1). For example, an EC C_i with ten users can satisfy 2-diversity property with SA's values ratio of $1 : 9$. Although, C_i is 2-diverse, but the SA values of someone's can be learned with 90% accuracy. The similar case is presented in Figure 6(b), where the SA's value, 'conservative' disclosure is 75% with accurate mapping of someone based on the QI's values. However, many variants of ℓ-diversity model have been proposed to solve these limitations. In addition, ℓ-diversity cannot be applied to the highly imbalanced (i.e., the data sets in which SA's values distributions is not uniform.) databases. In addition, ℓ-diversity privacy model degrades anonymous data utility significantly by not considering QIs distributions and similarities during anonymization.



	Quasi Identifiers		Sensitive Attribute
ID	Age	Country	Political views
1	35	Greenland	Liberal
2	35	Canada	Conservative
3	37	Canada	Conservative
4	37	Canada	Conservative

(**a**) Original data about the users (QIs and SA).

	Quasi Identifiers		Sensitive Attribute
ECs	Age	Country	Political views
C_1	35-37	North America	Liberal
	35-37	North America	Conservative
	35-37	North America	Conservative
	35-37	North America	Conservative

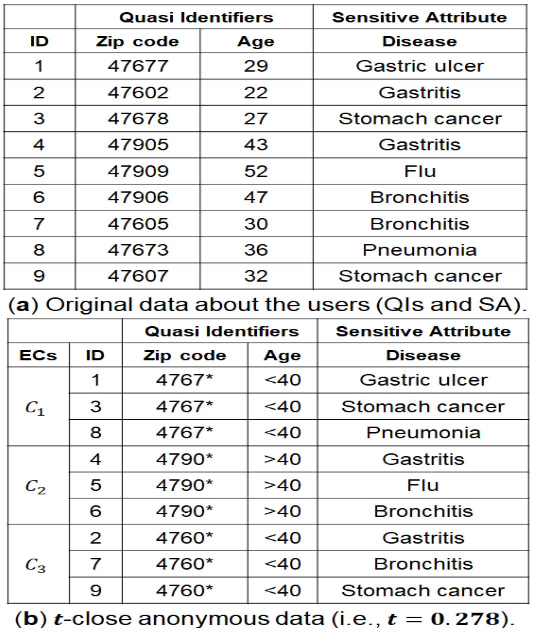
(**b**) 2-diverse data about the users (i.e., $\boldsymbol{l} = \mathbf{2}$).

FIGURE 6. 2-diversity applied to the user's relational data with four records.

C. *t*-CLOSENESS PRIVACY MODEL

The t-closeness privacy model [60] is a syntactic privacy approach used for the tabular data anonymization. It was proposed to solve the limitations of both k-anonymity and ℓ-diversity models in terms of privacy protection. A table T satisfies t-closeness if its records/tuples are split into ECs such that the distribution of SA in the whole T and the ECs of a t-close table T' are within t distance units of each other. The t-closeness privacy model suggests that the SA's values distribution in any EC of T' differs from the overall SA's values distribution in T by at most threshold t. The value of t can be determined by considering the protection level of the sensitive information and objectives of data publishing. The table shown in Figure 7(b) (adapted from [61]) is 0.278-close w.r.t disease SA. The table shown in Figure 7(b) is also 3-diverse (i.e., it contains three different values of SA in each EC.). The t-closeness privacy model significantly improves the user's privacy, but it severely reduces the utility of the released data. All of the above three privacy models are among the most famous syntactic privacy models, and many variants of these models have been proposed to resolve their limitations in the PPDP. In addition, many algorithms as an extension of these three privacy models have been proposed to combat with some specific types of threats that emerge from data sharing.

There exist many attacks that are possible, and can be easily launched on the T'. We can classify those attacks into two categories: (i) background knowledge attack, and (ii) flaws in an anonymization methods. In the former category of privacy attacks, the adversary has some known information about the target victim. For example, the QIs of an individual or some other information about existence of someone's data in a T'. The adversary uses such information to cause a privacy breach. In the latter category of privacy attacks, the flaws in an anonymization methods assist adversary in causing a privacy breach. For example, the homogeneity (i.e., no heterogeneity in SA's values in an EC) attack of the k-anonymity privacy model, skewness attack (i.e., no considerations of the SA's values' distribution in an EC) of the ℓ-diversity privacy model, and no semantic consideration (i.e., all disease information present in an EC belong to a same part/organ of a human body) in t-closeness privacy model lead to explicit privacy breaches in presence of the auxiliary information. All these attacks are practical, and can be launched on a T'. Hence, many improved variants of these three privacy models, and adversarial modeling based methods have been proposed to resolve these problems.



	Quasi Identifiers		Sensitive Attribute
ID	Zip code	Age	Disease
1	47677	29	Gastric ulcer
2	47602	22	Gastritis
3	47678	27	Stomach cancer
4	47905	43	Gastritis
5	47909	52	Flu
6	47906	47	Bronchitis
7	47605	30	Bronchitis
8	47673	36	Pneumonia
9	47607	32	Stomach cancer

(a) Original data about the users (QIs and SA).

		Quasi Identifiers		Sensitive Attribute
ECs	ID	Zip code	Age	Disease
C_1	1	4767*	<40	Gastric ulcer
	3	4767*	<40	Stomach cancer
	8	4767*	<40	Pneumonia
C_2	4	4790*	>40	Gastritis
	5	4790*	>40	Flu
	6	4790*	>40	Bronchitis
C_3	2	4760*	<40	Gastritis
	7	4760*	<40	Bronchitis
	9	4760*	<40	Stomach cancer

(b) t-close anonymous data (i.e., $t = 0.278$).

FIGURE 7. t-closeness (where $t = 0.278$) applied to the user's relational data with nine records.

D. DIFFERENTIAL PRIVACY MODEL

Differential privacy (DP) [18] is a well-known and mathematical definition-based privacy protection model. It is mostly used for privacy protection in an interactive settings of the PPDP. It protects the privacy of user by adding noise to the original user's data and it does not make assumptions about the intruder scenarios. The DP belongs to the semantic class of privacy models, and it yields superior privacy protection in PPDP compared to the syntactic privacy models. Considering the effectiveness of the DP model, U.S. census Bearu is planning to use the DP in their 2020 census, and all future data products [62]. It has been reported in the literature that DP provides a mathematically provable guarantee on privacy preservation against many privacy attacks such as differencing, linkage, and reconstruction attacks. DP concept can be defined as: given a dataset D_1, and a neighbour dataset D_2. Both data-sets differ in one and only one record, defined as $||D_1| - |D_2|| = 1$. In addition, a random function F with a range S, defined as $S \subseteq Range(F)$, satisfies DP if

$$Pr[F(x) \in S] \leq exp(\epsilon)Pr[F(y) \in S] + \delta \qquad (1)$$

In equation 1, variable $x, y \in D_1, D_2$; ϵ is a parameter; δ indicates a degree of the relaxation, and the probability is taking over the randomness of function F. In equation 1, if $\delta = 0$ then F satisfies ϵ-differentially private. In case of the query response (R_s), the probability Pr of DP model will be.

$$\frac{Pr(A(D_1) = R_s)}{Pr(A(D_2) = R_s)} \leq e^{\epsilon} \qquad (2)$$

In equation 2, A represents an anonymization algorithm, ϵ is a parameter, and its value is chosen by the data owners. Generally, the smaller value of ϵ is suitable for better privacy preservation in the PPDP.

In DP method, noise is added using the Laplace mechanism. DP mathematically ensures that any analyst/ data-miner seeing the result of a differentially private analysis will make the same conclusion about any individual's private information, whether or not that individual's private information is included in the input to the analysis. The overview of the DP is presented in Figure 8. Due to the strong privacy guarantees, the DP model has been extended by many studies for further improvements.

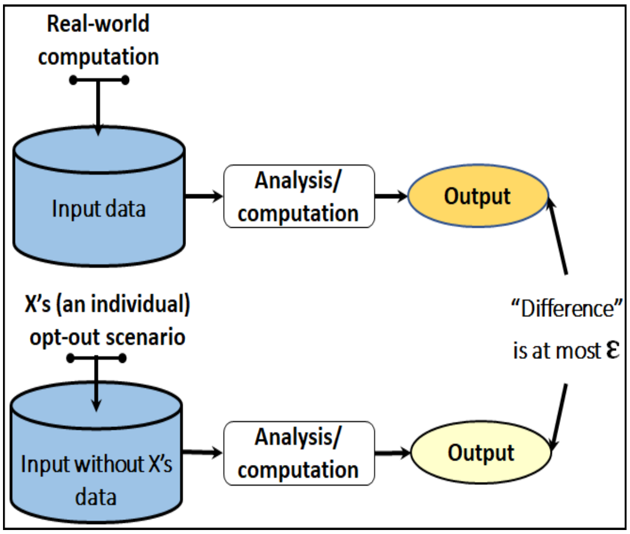


FIGURE 8. Functional overview of the differential privacy (DP) model used in the PPDP.

Many latest studies adopted DP concept for resolving the privacy issues in the both interactive and non-interactive settings of the PPDP. There exist numerous anonymization techniques which are ramifications of the four privacy models explained above. We summarize the approaches that extended the concepts of these four models in Table 1. We categorize the anonymization approaches into four main categories: (i) the k-anonymity model and its ramifications, (ii) the ℓ-diversity model and its ramifications, (iii) the t-closeness model and its ramifications, and (iv) the *DP* model and its ramifications.

We compare the existing approaches with each other based on six different parameters. The abbreviation used in Table 1 are: ID = identity disclosure, AD = attribute disclosure, MD = membership disclosure, IL = information loss, Sy. & Se. = syntactic and semantic, Bk = background knowledge, CoD = curse of dimensionality, Pr. = probability, SA = sensitive attribute, MSA = multiple sensitive attribute, LA = linkage attack, DMAR = discovering and maintaining association rules, DA = data analysis, ECs = equivalence classes, GPS = global positioning system, and QIs = quasi identifiers.

TABLE 1. Detailed comparisons of the different PPDP solutions used for relational data anonymization.

Privacy technique/model	Type	Restriction on the SA	SA's scenario	Anonymity operation on the QIs	Strength	Weakness
k-anonymity privacy model [12]	Syntactic	none	single	generalization/suppression	protection from the ID only in the PPDP	insufficient to protect from the AD
p-sensitive k-anonymity [63]	Syntactic	p distinct SA's values	single	generalization/suppression	protection against both ID and AD	high IL and computing complexity
p^+-sensitive k-anonymity [64]	Syntactic	p distinct SA categories	single	generalization	protection from both ID and AD	high utility loss on special purpose metrics
(α, k)-anonymity model [65]	Syntactic	α-threshold	single	generalization/suppression	protection against both ID and AD	high IL and QIs correlation loss
Complete (α, k)-anonymity [66]	Syntactic	α-threshold based protection	single	clustering	protection against both ID and AD	works only on the categorical data
$M3AR$ algorithm [67]	Syntactic	none	single	member migration	privacy protection by DMAR	AD is possible by not restricting SA's values
M-SA k-anonymity [68]	Syntactic	diversity	multiple	new records addition	better privacy and utility in the PPDP	may yield inaccurate analysis
Predictive delimiter algorithm [69]	Syntactic	dis-association	multiple	records shuffling	better privacy and accuracy	prone to the LA, and high complexity
HSVD algorithm [70]	Syntactic	equal distribution	multiple	generalization	protection against both ID and AD	high IL during analytics in most cases
Overlap slicing algorithm [71]	Syntactic	permute SA values	multiple	bucketization	reduces probability for adversaries to guess the SA values of users	low privacy in less diverse ECs
MSA(α, ℓ) algorithm [72]	Syntactic	(α, ℓ)-thresholds	multiple	clustering	privacy protection in the MSA cases	higher ID in presence of the BK
Cross-sampling algorithm [73]	Syntactic	perturbation	single & multiple	sampling	stronger privacy guarantees	explicit leakage of the SA's values
Flexible *GSC* algorithm [74]	Syntactic	permutation	single	Gray-code and binary conversion	privacy is equal or more than $1/k$	Can lead to sensitive patterns disclosures
ℓ-diversity privacy model [59]	Syntactic	ℓ-distinct SA's values	single	generalization/suppression	protection against the AD only.	insufficient to protect ID and MD
Independent ℓ-diversity [75]	Syntactic	ℓ-independent SA's value	single	perturbation and generalization	protection against the AD	lowest utility in terms of accuracy
LDSA algorithm [76]	Syntactic	semantic rules	single	generalization	effective privacy protection	lowest utility in terms of the IL
ℓ-cover Algorithm [77]	Syntactic	ℓ-diverse	single	safe generalizations	better privacy protection	missing experimental verification
(τ, ℓ)-diversity [78]	Syntactic	generalize SA values	single	ordering of the QIs	better privacy protection	work with categorical attributes only
Fast p-sensitive ℓ-diversity [79]	Syntactic	ℓ-diverse	multiple	clustering	protection against the AD	prone to the ID in presence of BK
Recursive (c, ℓ) diversity [80]	Syntactic	distinct ℓ-diverse SA's values	single	generalization	protects AD in the PPDP process	prone to explicit ID in imbalanced datasets
$DBTP - \ell - MDAV$ algorithm [81]	Syntactic	ℓ-distinct SA's values	single	micro-aggregation	yields higher data utility	prone to AD, ID, and MD in the PPDP
L_{sl}-diversity model [82]	Syntactic	ℓ-diversity group	multiple	grouping & suppression	greatly reduce the IL in micro data	run time is significantly higher
DI-Mondrian method [83]	Syntactic	ℓ-diverse	single	generalization	significant reduction in time and IL	it can lead to the SA disclosures
(ℓ, d)-semantic diversity [84]	Syntactic	ℓ-diverse with distance $\geq d$	single	dummy records addition	better privacy preservation for AD	cannot work well in the MSA scenarios
t-closeness privacy model [60]	Syntactic	t-close threshold	single	generalization/suppression	protects from both AD and MD	significant degradation in $T^{'}$'s utility
(n, t)-closeness model [85]	Syntactic	(n, t)-close threshold	single	generalization	protects from AD, ID, and MD	significant degradation in data utility
TCS algorithm [86]	Syntactic	t-close threshold	single & multiple	generalization	protects from MD,ID, and AD	prone to groups' identity theft problems
M-Shuffle algorithm [87]	Syntactic	SA grouping	single	QIs shuffling	excellent data utility and efficiency	explicit SA disclosure in imbalanced data
MSA t-closeness algorithm [88]	Syntactic	PCA and projections	multiple	clustering	privacy preservation in the MSA cases	privacy sensitive pattern leakage is possible
SABRE framework [89]	Syntactic	SA's redistribution	single	generalization	better privacy preservation	groups users' privacy invasion is possible
TCMA algorithm [90]	Syntactic	t-close threshold	single	micro aggregation	better privacy preservation	very high computational complexity
ASQI algorithm [91]	Syntactic	(ℓ, t)-thresholds	single	randomization	highest data utility in the PPDP	prone to AD and ID during published DA
TCMSA method [92]	Syntactic	t_X-close, where $o \leq t_X \leq 1$	multiple	generalization	better privacy protection in the MSA cases	poor data utility in imbalanced datasets
Differential privacy model [18]	Semantic	noise addition	single	QIs sampling	strong privacy guarantees in PPDP	significant degradation in anonymous data utility
(ϵ, σ)-differential privacy [93]	Semantic	noise addition	single & multiple	clustering	superior data utility for analysis	privacy breaches do occur with some Pr.
Extended ϵ-DP algorithm [94]	Semantic	noise addition	single	perturbation	higher data utility (i.e., accuracy)	lower privacy in presence of the BK.
Improved DP algorithm [95]	Semantic	noise addition	single & multiple	perturbation	better utility in interactive setting	no privacy guarantee for correlated data
Concentrated DP algorithm [96]	Sy. & Se.	random sampling of SA	single	random reshuffling	better privacy preservation in the PPDP	low utility in the sparse datasets
k-CRDP approach [97]	Semantic	noise addition	single	maximal information coefficient	better privacy protection of users	suffer from the CoD and high IL
MWEM algorithm [98]	Semantic	noise addition	single	as per query budget	better user's privacy preservation	work well with only categorical attributes
DPMK anonymity [99]	Se. & Sy.	noise addition	single	generalization	superior anonymous data utility	very high computing complexity
IPA Method [100]	Semantic	noise addition	single	generalization, insertion, and suppression	better data utility for medical researchers	may yield privacy disclosures in presence of strong BK and auxiliary information
Pareto Principle-based DP [101]	Semantic	noise addition	single & multiple	DP with Pareto principle	improve accuracy of the query results	works with only limited type of queries
ϵ-DP based privacy protection [102]	Semantic	distort and reconstruct	single	ϵ-DP with trajectory data shuffles	protects sensitive places visits of user	poor utility with raw GPS data
(ϵ,δ)-local DP [103]	Semantic	correlated noise addition	single	synthetic datasets creation	higher accuracy using SVM classifier	poor utility in general purpose metrics

Some studies have used the combinations of more than one privacy models (e.g., k-anonymity and ℓ-diversity) to protect the user's privacy in data publishing. Majeed *et al.* [104] extended the k-anonymity model to effectively resolve the privacy and utility trade-off in the PPDP. The proposed scheme performs adaptive data generalization considering both the vulnerability of the QIs and the diversity of the SA to anonymize data. Jordi *et al.* [105] suggested that t-closeness can be extended to the DP model when $t = exp(\epsilon)$. The authors proposed a method for achieving both t-closeness and DP, respectively. As a result, higher utility was retained in the anonymous data compared to the noise addition methods. Rasool *et al.* [106] used the k-anonymity concept in the clustering process to produce anonymous data set for publishing. The proposed *SBC* algorithm significantly reduces information loss and retains better semantics of the original dataset. Recently, the k-anonymity concept has been extended in combination with entropy concept to protect the users' groups privacy in data publishing [107]. The proposed anonymization method effectively resolves the users' groups privacy issues stemming from the low diverse ECs, and highly susceptible QIs present in a person-specific dataset.

E. METRICS USED FOR THE EVALUATION OF PRIVACY AND UTILITY OF RELATIONAL DATA ANONYMIZATION TECHNIQUES

1) PRIVACY EVALUATION METRICS

There exist multiple ways to quantify the privacy protection offered by an anonymization algorithm. The five common methods that are employed to evaluate the effectiveness of any anonymization algorithm in terms of privacy protection are: (i) anonymous and original dataset linking and calculating the probability of successful matches based on the QI values in both these datasets. The probability value tells the amount of privacy protection an anonymization algorithm will offer during published data analysis. In this case, it is assumed that attackers may have access to the excessive amount of auxiliary information from some external sources (i.e., voter list, online repositories, e-commerce sites etc.) to launch identity and attribute disclosure attacks; (ii) privacy protection evaluation in presence of the background knowledge. In this privacy evaluation metric, the data owner assumes that attacker may possess the true information (i.e., age and gender of a Bob) about some users and he/she can explore only the particular ECs to infer private information of relevant user/users. To evaluate effectiveness of algorithm in this regard, the data owner can pick some instances from the original data and can evaluate the anonymization algorithm privacy protection level by matching; (iii) privacy protection evaluation with the help of privacy-sensitive (PS) rules. In this case, the data owner can construct certain rules to evaluate privacy protection. For example, how many people having age >40 suffer from this disease(i.e., cancer). The sensitive knowledge pattern revelation, and attribute and identity disclosure of multiple users through PS rules have a wide range of negative consequences on people's life; (iv) prediction about the users SA through the existence of private and public profiles in SNs. In this case, the data owner can quantify the amount of protection level of an algorithm assuming that either partial or full user's original data is known to the attacker as some users willingly publish their QIs over different SNs. The factual information that can be learned through the knowledge gained from the D to invade unknown users' privacy can be used to evaluate an anonymization effectiveness; (v) privacy protection evaluation in the presence of malicious users' in a dataset. In this case, the data owner can classify some of the tuples as malicious and can calculate the similarity with other (i.e., non-malicious) users to quantify the privacy protection level. In some cases, the sensitive queries and corresponding private information budget is also used for the evaluation of anonymization algorithms/models.

2) UTILITY EVALUATION METRICS

During tabular data anonymization, the original QI's values are modified to fulfill the privacy needs, hence, the data utility degrades. There exist multiple ways to quantify the anonymous data utility offered by an anonymization algorithm. We classify the metrics used for measuring anonymous data utility into two categories: special purpose and general purpose metrics. The special purpose metrics use machine learning methods to measure the anonymous data quality. The most widely used special purpose metrics are, accuracy or error rate, F-measures, precision, and recall. The general purpose metrics measure the information loss caused by modifying the original data. The most popular general purpose utility evaluation methods are, weighted certainty penalty, generalized information loss (*GenILoss*), discernability metric, minimal distortions, average equivalence class size (C_{AVG}), KL-divergence, granularity, query accuracy, global loss penalty (GLP), normalized mutual information (NMI), relative error (RE), and information theocratic metrics (ITM). The comprehensive details about these utility metrics can be found in the recent studies [5], [9], [35], [50].

F. PRIVACY PRESERVING DYNAMIC DATA PUBLICATION

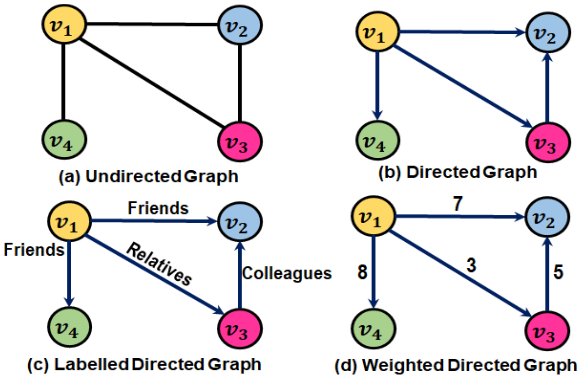
Privacy preserving dynamic data publication (PPDDP) allows organizations to publish a dataset multiple times. PPDDP enables organizations to share up-to-date data with multiple recipients statically or after some modifications (i.e., update, delete or insert). In contrast, privacy preserving static data publication (PPSDP) focused only on one time publication of a dataset, and many approaches have been proposed for the PPSDP in literature [12], [59], [60]. Meanwhile, PPDDP opens a new era in PPDP research, and many organizations are publishing their users data dynamically. Kabou *et al.* [108] presented a comprehensive survey about the PPDDP, and summarized different studies used for the PPDDP. Shi *et al.* [109] presented a method for PPDDP using distance and information entropy concepts. The proposed method ensures that individual privacy is preserved after the data has been subjected to multiple releases. The m-invariance privacy model [110] was proposed to limit the risk of privacy disclosure in data re-publication. The proposed approach jointly uses the m-invariance and counterfeited generalization concepts to solve the PPDDP problem. Anjum *et al.* [111] proposed a τ-safety privacy model for the PPDDP. The proposed approach performs better in the presence of external and internal updates. Zhu *et al.* [112] proposed a τ-safe (ℓ, k)-diversity privacy model for sequential publication. The proposed privacy model guarantees that each record's signatures/values keep consistency or have no intersection in all data releases. This model can be applied to the data in which individual has multiple records. The proposed algorithm performs better compared to the m-invariance and τ-safety privacy models. Considering the widespread applications/uses of the privacy preserved published data, the PPDDP has become an emerging area of research in recent years.

G. CHALLENGES IN THE RELATIONAL DATA ANONYMIZATION

The representation of tabular data is relatively simpler than the SN data. In relational data, each row/tuple represents one real world entity/individual. We identify following five challenges that make the relational data anonymization challenging. First, selection of the user's attributes that are regarded as QIs. In relational data, a small subset of the users' attributes are chosen as QIs. However, some QIs can behave as SA (i.e., profession) in practice. Thus, appropriate selection of QIs prior to data sanitization is imperative. Second, over reliance on the custom made QI's values generalization taxonomies. The relational anonymization is performed with the help of pre-defined taxonomies of the QIs. Meanwhile, these taxonomies do not truly reflect the privacy and utility trade-off. Thus, devising the appropriate taxonomies with in-depth analysis of QI's domain values in relational anonymization is challenging. Third, tabular data anonymization in multiple SAs (MSAs) scenarios. In some cases, the tabular data contains multiple SAs about individuals. Anonymizing the tabular data having MSAs is harder compared to the single SA scenarios due to the high risk of user's private information disclosures. Fourth, quantifying the impact of QIs on both user's privacy and anonymous data utility. Since each QI affects the privacy and utility differently, and some QIs can be highly vulnerable in terms of privacy and some QIs have higher utility. Thus, quantifying each QI's statistics related to privacy and utility is very challenging in the PPDP. Fifth, accurate estimation of the subjects' re-identification risk. The existing PPDP methods often assume worst-case scenarios regarding attackers, and apply heavy changes in the data that impact the shared data utility and often limit data sharing on a wider scale. Thus, accurate and novel adversarial modeling methods are needed to support PPDP in big data era. Aside from the five potential challenge explained above, in some cases, an adversary can possibly link the whole published dataset (e.g., T') by leveraging the auxiliary information [113]. Hence, devising algorithms/method which can estimate the users' privacy disclosure risk as accurate as possible during published data analytics is a vibrant area of research.

V. STRUCTURAL ANONYMIZATION TECHNIQUES USED FOR THE SOCIAL NETWORKS DATA ANONYMIZATION

Structural anonymization refers to the modification in the structural properties of the social network (SN) data (i.e., graphs) to protect the privacy threats that emerge from SN data publishing. Generally, the SN analysts represent the SN data mainly via two methods: metrics and graphs [114]. The matrices representation of the SN data allow the application of computer tools and mathematical models to summarize and extract patterns. Since finding relevant patterns from the dense graphs is extremely complex. The adjacency matrix is used as the variant of the graphs in social network analysis. For instance, a G with n users can be modeled as an adjacency matrix M of size $n \times n$. In an adjacency matrix, the relationship between two users i and j can be represented by the value ($< 0, 1 >$or $< y, n >$) in a cell i, j. To represent various forms of users' data and to model the structural properties of SNs, a G can have their nodes and edges labeled or unlabeled, undirected or directed, weighted or unweighted as presented in Figure 9.



(a) Undirected Graph

(b) Directed Graph

(c) Labelled Directed Graph

(d) Weighted Directed Graph

V	v_1	v_2	v_3	v_4
v_1	0	1	1	1
v_2	1	0	1	0
v_3	1	1	0	0
v_4	1	0	0	0

(e) Social Network representation using a matrix with four users.

FIGURE 9. Overview of the social network representation using different forms of graphs and a matrix.

Due to the inefficacy in handling the complex SN data, and inability to correctly represent the SN with users' attribute information, matrices are rarely used to represent the SN data. In contrast, graphs can represent the users' attribute information properly along with social relations. Therefore, graphs are most widely used in SN analysis, and assist effectively in proving graph-based theorems. This work uses SN data modeled as a graph for further analysis and discussions. Users' privacy preservation in SN data publishing is very challenging compared to the relational data due to the more

TABLE 2. Description about the pieces of information related to user's privacy in SNs data (Ref. [39]).

Sr. No.	Pieces of information	Description about each piece of information
1	Node existence/non-existence	Target individual presence/absence in a SN data (i.e., graph).
2	Node properties	Some properties of a node such as degree and position/placement etc. in a graph.
3	Link relationship	The link relationship (e.g., patient & HIV specialist) among nodes in a SN data.
4	Sensitive node label	The sensitive label (i.e., disease information) carried by an individual in a SN data.
5	Sensitive edge label	The sensitive edge label (i.e., social connection) between vertices in a SN data.
6	Graph metrics	The sensitive graph properties (i.e., betweenness, centrality, closeness, and reach-ability) in a SN data.
7	Link weight	The edge weight (i.e., communication cost) between nodes in a SN data.
8	Node-group affiliation	The link relationship between nodes (i.e., users) and online SN groups in a SN data.

TABLE 3. Types of the background knowledge (BK) used by the adversaries to jeopardize user's privacy in the PPGP.

Sr. No.	Type of the BK	Examples	Privacy threats that emerge from the background knowledge
1	Node/vertex degrees	Number of friends	Identity disclosure and membership disclosure
2	Attributes of nodes	Sex, age, occupation, disease	Identity disclosure and content disclosure
3	Link relationship	Communication channel, social connection, online group	Content disclosure and affiliation disclosure
4	Neighborhoods	Friend's circle, close friends, mutual friends	Identity disclosure and content disclosure (with some probability).
5	Neighborhood-pair properties	Close friends, mutual friends, friends of friends (FoF)	Identity disclosure with very high probability.
6	Embedded sub graphs	Vertices, edges, clique	Identity disclosure and groups identity theft.
7	Knowledge graph	Sex, age, relations, followers, following	Identity disclosure and content disclosure.
8	Accounts on multiple SNs	Profile information, login details, display name	Identity disclosure and content disclosure.
9	Graph metrics	degree, centrality, closeness, sub graphs	Identity disclosure, content disclosure, and groups identity theft.
10	User behaviors	recommendation, interests, SN use patterns	Identity disclosure and content disclosure.
11	Auxiliary information	Other SNs graphs	All four disclosures about an individual, and groups identity/attribute theft.

pieces of private information contained in a G. The pieces of information concerning user privacy in a social graph data are summarized in Table 2. The SN users need privacy protection for most of the pieces of their private information shown in Table 2. In contrast, the relational data mainly contains four pieces of information about the users: (i) users' QIs, (ii) users' SA, (iii) micro-statistics (i.e., SA's particular value shared by less number of people in a dataset) about users, and (iv) macro statistics (i.e., SA's particular value shared by large number of people in a dataset) about the users. In addition, due to the availability of user's profiles information on the SNs sites and accounts on multiple SNs, the SN users privacy can be compromised easily compared to the tabular data.

The background knowledge (BK) is the fact or an information known to the adversaries about an individual or group of individuals, which can be exploited to infer the SA of an individual(s) from the G'. The BK can be acquired from different sources, and its degree purely depends upon the adversaries' capabilities and technical knowledge. In practice, it is very difficult to quantify the level of the BK possessed by the adversaries, and many existing algorithms assume certain pieces of information as BK while anonymizing user's data. The BK types with sufficient details and examples are explained in literature [37], [39], [41]. In Table 3, we summarize the most recent types of the BK that are used by the adversaries to jeopardize SN user's privacy during published graphs analytics.

The most common technique used for SN users privacy preservation in the PPGP is anonymization. After in-depth synthesis of the literature [114]–[116], we present the taxonomy of the PPGP approaches along with representative anonymization methods employed for SN data in Figure 10. These PPGP approaches can be broadly classified into five categories, namely graph modification techniques, graph generalization/clustering techniques, privacy aware graph computation techniques, differential privacy based graph anonymity techniques, and hybrid anonymization techniques. Brief description along with relevant examples about all five anonymity techniques used in the PPGP is given in subsequent paragraphs.

Due to the widespread applications of the SNs data, many structural anonymization techniques have been proposed for the PPGP. These techniques modify the structure of the SN graph by adding/ deleting vertices or edges to preserve the user's privacy. Aside from the add/delete, in some case, edges and vertices are switched or re-arranged in clusters to preserve user's privacy. The overview of anonymous graphs obtained by adding vertices and edges is shown in Figure 11. In Figure 11(b), two new edges have been created ($\{v_1, v_3\}$ and $\{v_2, v_4\}$). Similarly, in Figure 11(c), two new nodes ($\{v_7, v_8\}$) with four edges ($\{v_7, v_3\}$, $\{v_7, v_2\}$, $\{v_8, v_1\}$ and $\{v_8, v_2\}$) have been added in the anonymous version of a G.

The addition/deletion of the nodes and edges can be constrained or random (e.g., non-constrained) depending upon the scenario.

An example of the original graph G anonymization through both constrained and random perturbation techniques taken from study [30] is shown in Figure 12(i), (ii). Figure 12(i)(b) shows an example of a perturbed version of the network shown in Figure 12(i)(a) by Rand add/del operation. In this example, the two edges ($\{v_1, v_5\}$ and $\{v_2, v_3\}$) have been removed and two new edges ($\{v_6, v_7\}$ and $\{v_8, v_9\}$) have been added to produce the anonymous graph G'. Meanwhile,

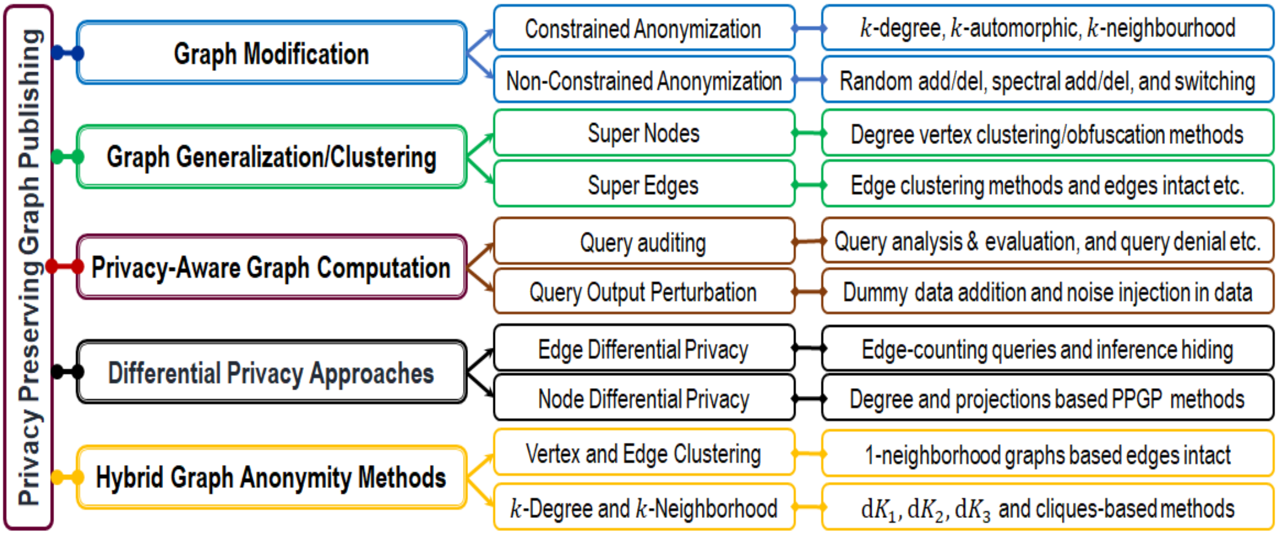

Privacy Preserving Graph Publishing
Graph Modification
Constrained Anonymization
k-degree, k-automorphic, k-neighbourhood
Non-Constrained Anonymization
Random add/del, spectral add/del, and switching
Graph Generalization/Clustering
Super Nodes
Degree vertex clustering/obfuscation methods
Super Edges
Edge clustering methods and edges intact etc.
Privacy-Aware Graph Computation
Query auditing
Query analysis & evaluation, and query denial etc.
Query Output Perturbation
Dummy data addition and noise injection in data
Differential Privacy Approaches
Edge Differential Privacy
Edge-counting queries and inference hiding
Node Differential Privacy
Degree and projections based PPGP methods
Hybrid Graph Anonymity Methods
Vertex and Edge Clustering
1-neighborhood graphs based edges intact
k-Degree and k-Neighborhood
dK_1, dK_2, dK_3 and cliques-based methods

FIGURE 10. Taxonomy of privacy preserving graph publishing (PPGP) approaches used for SN data.

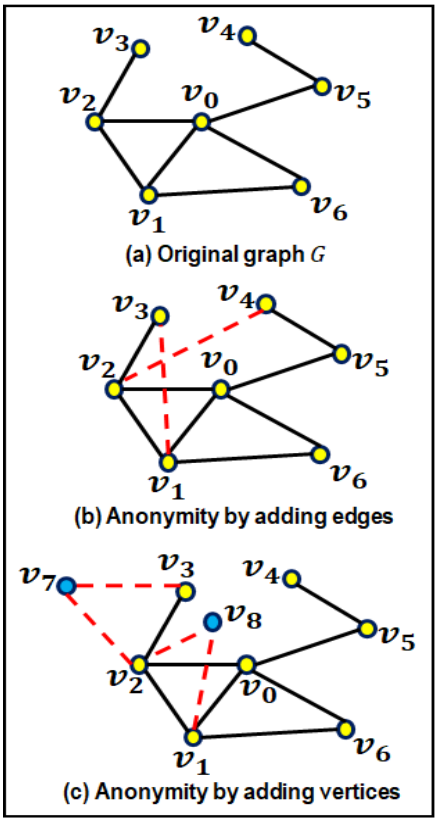


FIGURE 11. Examples of anonymizing graph by adding edges and vertices.

the perturbed version of the graph shown in Figure 12(i)(c) is obtained from the Rand switch operation. In this example, two edges ($\{v_1, v_2\}$ and $\{v_4, v_5\}$) were switched to ($\{v_1, v_4\}$ and $\{v_2, v_5\}$) to produce the anonymous graph G'. In the random perturbation (e.g., non-constrained), there is no hard constraints regarding the edges addition/deletion/switching.

In contrast, in the constrained anonymization, the nodes/edges addition/deletion is bounded by some constrains (i.e., degree). Figure 12(ii)(b) shows an example of the perturbed graph G' obtained by applying edge modification concept on an original graph G given in Figure 12(ii)(a). The perturbed graph is k-degree anonymous, where $k = 2$. The original graph G has a degree sequence $d(G) = \{2, 4, 2, 1, 3, 2, 2, 2, 2\}$, while the modified graph G' given in Figure 12(ii)(b) has degree sequence $d(G') = \{2, 3, 2, 2, 3, 2, 2, 2, 2\}$. The number of vertices and edges are same in both graphs, and the anonymous graph is 2-degree anonymous (i.e., each vertex has at least 2-edges). The single modification in a G by modifying edge ($\{v_3, v_2\}$ to $\{v_3, v_4\}$) has made the G two anonymous. The constrains value can be adjusted considering the protection level and graph structure. Another example of the 2-degree anonymous G' by adding two edges ($\{v_4, v_{10}\}$ and $\{v_5, v_{10}\}$), and one vertex (v_{10}) is shown in Figure 12(ii)(c). The degree sequence of G' becomes $d(G') = \{2, 4, 2, 2, 4, 2, 2, 2, 2, 2\}$ and number of vertices and edges increase by one and two, respectively. In the constrained perturbation, addition/deletion of the nodes/edges follow some criteria (i.e., degree, closeness, and clustering co-efficient etc.), and further addition/deletion of the nodes/edges stop once the defined criteria is satisfied.

There are six basic edge and vertex modifications techniques for SN data anonymization [44]. The modifications techniques are: (i) edge add, (ii) edge delete, (iii) edge add/del, (iv) simple edge switch, (v) double edge switch, and (vi) node addition. Most of the existing SN data anonymization methods use one (or more) of these six modifications techniques during graph anonymization. A detailed taxonomy of the graph modification techniques is given in study [53]. The four types of graphs that are mainly used to represent the SNs users' data are: simple graph, bipartite graph, labelled graph, and uncertain graph.

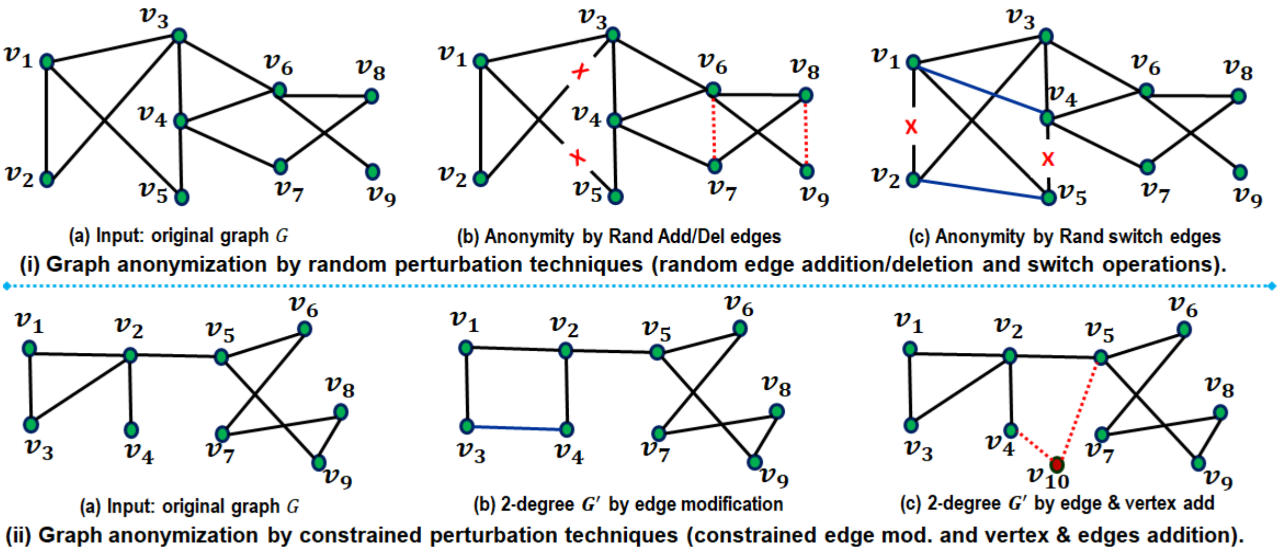

v_1
v_2
v_3
v_4
v_5
v_6
v_7
v_8
v_9
v_{10}
(a) Input: original graph G
(b) Anonymity by Rand Add/Del edges
(c) Anonymity by Rand switch edges
(i) Graph anonymization by random perturbation techniques (random edge addition/deletion and switch operations).
(a) Input: original graph G
(b) 2-degree G' by edge modification
(c) 2-degree G' by edge & vertex add
(ii) Graph anonymization by constrained perturbation techniques (constrained edge mod. and vertex & edges addition).

FIGURE 12. Original graph anonymization by using random (e.g., non-constrained) and constrained anonymization methods.

In SN data anonymization, majority of the approaches are driven from the concepts that were proposed for anonymizing tabular data. For example, the k-anonymity model and its variants such as k-degree anonymity, k-isomorphism anonymity, k-automorphism anonymity, k-candidate anonymity, k-neighborhood anonymity, and (k, ℓ)-grouping have been adapted to anonymize SN data. The generalization/clustering based approaches anonymize SN data by partitioning it into different clusters, and generalizing the clusters into super nodes/edges [117]. The concept of these approaches after clustering is analogous to the EC generalization of the relational data. Moreover, the cluster sizes and generalization degrees are determined in a way that maximal information is retained in the clustered network (i.e., G'). The conceptual overview of the generalization/clustering based anonymization is presented in Figure 13.

A network/graph with seven nodes and two QIs (age, gender) is provided as an input (Figure 13 (a)), whole network is partitioned into three clusters (c_1, c_2, c_3) based on the QI's similarities (Figure 13 (b)), and a corresponding generalized network with three super nodes is obtained as an output (Figure 13 (c)). We used three distinct symbols (e.g., square, circle, and triangle) to denote the users in each cluster and super nodes, respectively. The two numbers in each super node represent the cluster size (e.g., number of users) and intra-cluster edges. For example, in cluster c_3 there are two users but there is no edge between them. Therefore, the value inside the triangle (a.k.a super node) is (2, 0). The weighted edges between super nodes represent the inter-cluster edges.

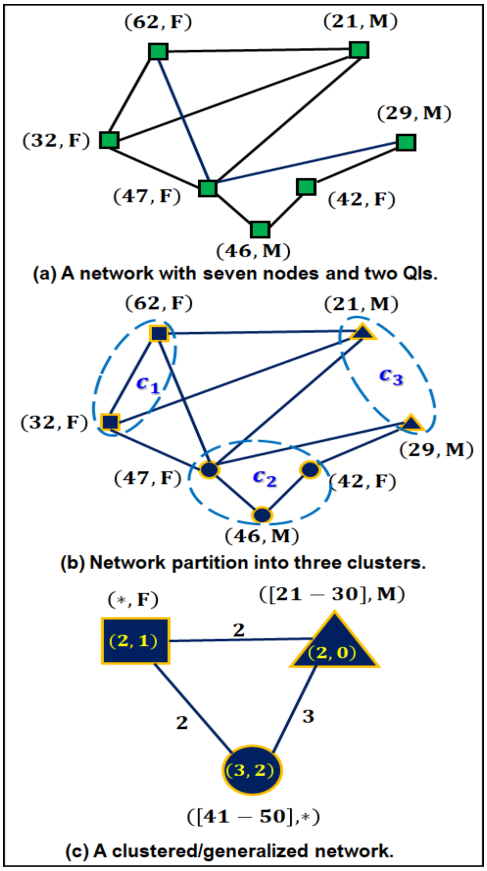
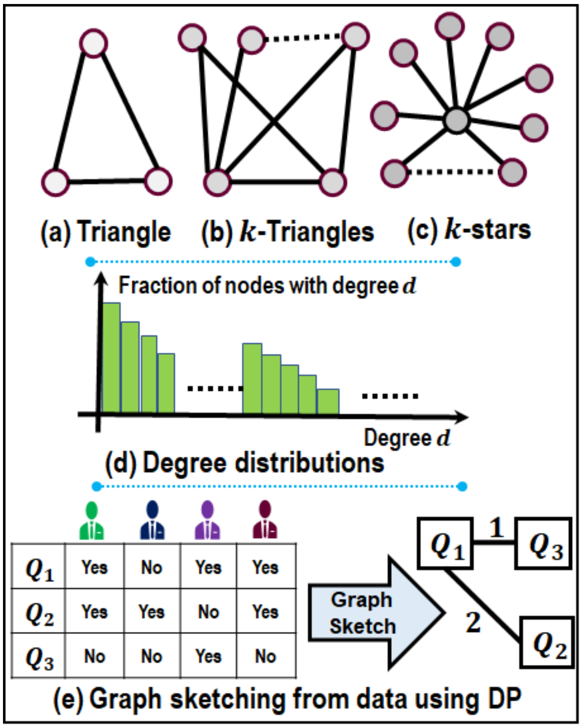


FIGURE 13. Original graph anonymization by employing clustering/generalization based method.

On the contrary, the privacy-aware graph computation and DP based approaches do not release the entire G' like previous two techniques (e.g., graph modification and graph generalization/clustering techniques) [30]. These approaches perform computations on the original graphs G, and yield output of an analysis computation. Compared to the previous two techniques, these approaches allow constrained analysis on the SNs data that can limit the widest range of applications for knowledge extraction and data mining. The DP concepts have been tailored with the structural properties to anonymize graphs data, and DP is regarded as one of the best privacy-aware graph computation techniques [30]. The DP based approaches in SN data anonymization have been classified into three categories, node-level DP, edge-level DP, and node and edge level DP. These techniques compute the useful statistics from an original graph in such a way that individual's privacy is preserved, and data remains useful for the analytical purposes.

The useful analysis provided by privacy-aware graph computation and DP based approaches are: graph density, edges count, relationships degree, degree distributions, size of the network, centralities, closeness, counts of the sub graphs, top k-users with highest degree in a network, distance/similarity between users, path length, clustering coefficients, community discovery/extraction, hypergraphs, joint degree distribution, cuts, number of users with degree d, aggregation, projections, and sparse and dense segments of the graphs, to name a few. These valuable statistics can be utilized for range of applications including social network analysis, marketing, preference mining and analysis, collaborative filtering, epidemiological investigation, and information spread/contagion etc. A sample of the graphs' statistics computed with the help of node DP techniques [118], [119] and minimum spanning tree (MST) DP [120] approaches is presented in Figure 14.

Aside from these basic statistics, applying node/edge DP approaches for publishing graph cuts and pair-wise distance between nodes are handy for data-driven applications. Furthermore, many ℓ-diversity and t-closeness variants have been proposed by the researchers to solve the content disclosure problems in the PPGP. A comprehensive details about these variants with definition is elaborated in studies [33], [55], [121]. Despite the success of existing PPGP mechanisms, preserving privacy in dynamic SNs is still an important research direction which has received significant attention from the research community recently.



Q_1	Yes	No	Yes	Yes
Q_2	Yes	Yes	No	Yes
Q_3	No	No	Yes	No

FIGURE 14. Overview of the graphs' statistics determined by privacy-aware graph computation methods.

The hybrid anonymity approaches usually employ more than one anonymity technique to yield anonymized SN data [122]. However, the complexity of the hybrid anonymity approaches is relatively higher when G contains substantial number of nodes and edges. Hence, hybrid anonymity techniques are used only in specific scenarios involving SN data. We summarize the recent structural anonymization approaches used for SN data (e.g., graphs) in Table 4. The abbreviation used in Table 4 are: EM = edge modifications, PPGP = privacy preserving graph publishing, PUT = privacy-utility trade-off, CC = computational complexity, SNs = social networks, SA = sensitive attributes, Bk = background knowledge, GP = graph publishing, GA = graph anonymization, EC = equi-cardinal, and IL = information loss. Just like relational data, many privacy and utility evaluation metrics have been proposed to access the performance of SNs data anonymization mechanisms.

A. METRICS USED FOR THE EVALUATION OF PRIVACY AND UTILITY OF STRUCTURAL ANONYMIZATION TECHNIQUES

1) GRAPHS UTILITY EVALUATION METRICS

The existing well-known graph utility evaluation metrics are: (1) degree (Deg.), (2) effective diameter (ED), (3) joint degree (JD), (4) local clustering co-efficient (LCC), (5) path length (PL), (6) closeness centrality (CC), (7) eigen vector (EV), (8) betweenness centrality (BC), (9) network constraints (NC), (10) network resilience (NR), (11) infectiousness (Infe.), (12) page rank (PR), (13) hub score (HS), (14) global clustering co-efficient (GCC) or global transitivity (GT), and (15) authority score (AS). Aside from these well-known metrics, some general purpose approaches such as accuracy, classification, information loss (IL), ratio of top influential users (RRTI), query response, and graph spectral properties have also been used to measure anonymous graph's utility. Furthermore, there exist seven application-specific utility evaluation methods which are, (1) role extraction (RX),(2) reliable email (RE), (3) minimum sized influential nodes set (MINS), (4) influence maximization (IM),(5) community detection (CD), (6) source routing (SR), and (7) sybil detection (SD). We refer interested reader to the previous work [123] for more detailed definitions and descriptions of the above metrics.

2) GRAPHS PRIVACY EVALUATION METRICS

The most commonly used privacy metrics in evaluating the performance of graph anonymization algorithms are, (1) the number of re-identified nodes, (2) degree of uncertainty, (3) adversary's success rate, (4) query's error, (5) information gain, (6) amount of leaked information, (7) information

TABLE 4. Detailed comparison of recent structural anonymization approaches used for the social networks data (i.e., graphs data) anonymization.

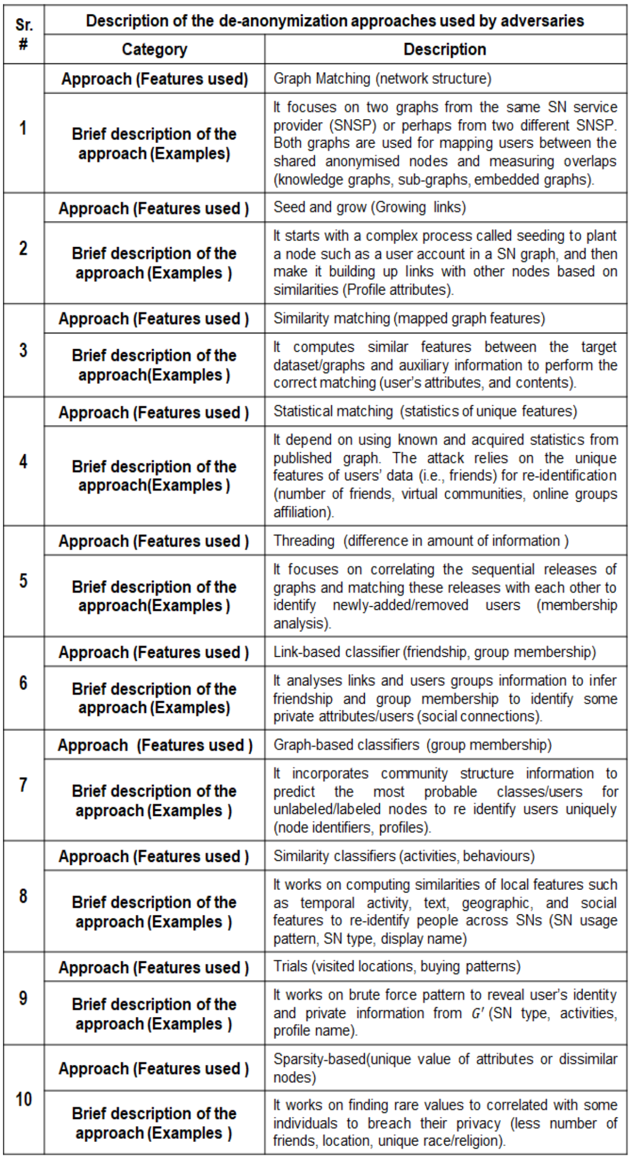
Method [Ref.]	Assertion (e.g., problem solved)	Anonymization by	Merits	Demerits
k-degree anonymity [28]	preserves identity and link disclosures	edge additions	scalable to the large SN graphs	high IL on sparse and complex graphs
KDVEM:k-degree anonymity [126]	preserve identity disclosure	modifying edges and vertices	less changes in the graph's structure	poor utility in large SN graphs
k-degree sequence anonymity [127]	preserve identity disclosure	modifying edges	better SN data utility in the PPGP	complexity rapidly increases with the k
k-anonymity on graphs [128]	preserve identity disclosure	adding dummy vertices and EM	better utility on three graph metrics	works well on small graphs only
structural integrity [129]	preserve identity disclosure	edges addition and deletion	improved graph's utility for analysis	very high computational cost
randomization [130]	preserve link disclosure	edge shifting	resolves privacy and utility trade-off in PPGP	identity disclosure is possible due to the BK
rough set based approach [115]	protects sensitive relationship disclosure	edge shuffling	improved graph utility for analytics	vertex re-identification possible in presence of BK
randomized perturbations [131]	preserve identity disclosure	modifying edges and vertices	better graph data utility for analytics	prone to the content disclosure threats/problems
(k,ℓ)-anonymity [132]	preserve identity and content disclosures	edge addition	produces more useful graphs	high computational cost on the large graphs
SN immunization (SNI) [133]	preserve content disclosure	guarding nodes	solves multiple SAs problem in the PPGP	lower utility in terms of the IL
EC clustering [134]	preserve identity, link, and content disclosures	vertex shuffling	better privacy for all three disclosures	lower utility on the complex graphs
PM index [135]	preserve identity disclosure	perturb links	better users' privacy protection	lower utility and very high computational complexity
$(k,\tau_{G,\ell})$-adjacency [136]	preserve identity disclosure	edge set perturbations	less graph edits in the GA	poor utility on the dense graphs
mathematical modeling [137]	preserve identity disclosure	edge addition	better privacy preservation during $G^{'}$ analysis	less utility on the sparse graphs
active learning method [138]	preserves link predictability	deleting edges	protects from the inference attacks	may deletes highly useful links/edges
edge entropy [139]	preserve identity disclosure	edge modification	better privacy of only some nodes	unable to resolve PUT in the PPGP
safely permuting condition [121]	users, groups and edges' privacy protection	nodes masking	suitable for dynamic GP	high complexity and low utility in terms of analysis
Anatomy based clustering [140]	preserves identity, link and attribute threats	clustering	better utility of the SN graphs	high computing cost and more changes in G
statistical analysis [141]	preserves identity disclosure	edge removal	better privacy in the PPGP	higher IL due to more edits in G
Bayes rule (BR) [142]	preserves identity disclosure	edge addition	better PPGP method in practice	does not provide privacy analysis on the large graphs
clustering perturbation [143]	preserves vertices properties	modifying edges and exchanging attributes randomly	better privacy preservation in the PPGP	lower utility in terms of the IL
k-degree-l-diversity [144]	preserves identity and content disclosures	adding noise nodes	superior privacy protection in the GP	lower data utility and high CC
k-subgraph [145]	preserves identity disclosure	move edges to achieve k-degree	defense capability against malicious attacks	limited testing was carried out
k-degree anonymity [146]	preserves identity disclosure	adding noise nodes	slight modification in the original graph	lower utility due to more noise addition
partial k-anonymity [147]	preserves identity disclosure	adding and deleting edges	better utilities in the PPGP's analytics	privacy breaches occur in presence of the BK
$wPINQ$ method [148]	preserves identity disclosure	adding new nodes	better data utility during published graph analysis	higher modifications in the G's structure
$CDGA$ method [149]	preserve identity disclosure	remove edges	slight changes in the original graph	prone to link disclosure threats
DP-FT method [150]	preserve identity and link disclosures	masking vertices	maintains good utility during the GA	prone to content disclosure threats
$HDP+AUL$ method [151]	preserve link disclosure	edges modification	maintains good privacy in the PPGP	less utility due to noise addition
Differential privacy [152]	preserve link disclosure	edges modification	less changes in the original graph	errors rate is still relatively high in $G^{'}$
$EN-DP$ method [153]	preserve identity disclosure	ordered edge insertion	effective in terms of less changes in the G	more utility loss on the large G
$dK-1,2,3$ [154]	preserve identity and link disclosures	nodes and edges perturbation	more utility in terms of query response	more changes in the graph structure
subgraph-DP [155]	preserve groups identity disclosure	edges addition	slight changes in the graph	no guarantee for single user's privacy
HIGA method [156]	preserved identity disclosure	edge addition and deletion	better utility in terms of query's answer accuracy	prone to content and link breaches
graph anonymity [157]	preserved link disclosure	edges modification	protects from the inference attacks	high utility loss due to more changes in G
k-decomposition [158]	preserved identity disclosure	graph modifications	better PUT resolution in the GP	poor scalability on large graphs
random walk [159]	preserved link disclosure	degree and link perturbations	hides both hub nodes and social relationships in SNs	high modifications in the G's structure
(k,ℓ)-anonymity [160]	preserved link and identity disclosures	edge additions	protection from the active attacks	cannot thwart the sybil retrieval
Differential Privacy [161]	preserved link and content disclosures	node sampling	better user's privacy protection	very high computing cost
data-manipulating methods [162]	preserve content disclosure	indistinguishable links	lowers prediction accuracy for the SA	utility can be lower on large graphs
heuristic PBCP method [163]	preserved content disclosure	clustering	better privacy protection in the GP	prone to the link disclosure threats
$KDDLD-UL$ [164]	preserved content and identity disclosures	nodes addition	better utility and low computing cost	prone to the link disclosure threats
(k,ℓ)-grouping [165]	preserve affiliation link disclosure	generalizes node's attributes	better privacy protection in the PPGP	more changes in the G's structure
PBCN approach [166]	protects from graph structure and degree attacks	clustering and noise addition	resolve "trade-off" between data utility and privacy	more parameters and high computing complexity
DPCD and NAR algorithm [167]	protects user's private social relationships and attributes	DP-based community detection and NAR algorithm	privacy preservation of both social relationships and attributes	may lead to privacy breaches in dense structures G
SSN algorithm [168]	protects the privacy of structural role	variational Bayes-weighted network DP (*VB-WNDP*)	significant improvements in anonymous graph accuracy	limited experiments were performed for validation

surprisal, (8) privacy score, (9) association rule hiding (ARH), (10) distribution leakage, (11) prior information belief and posterior information belief, (12) entropy leakage, (13) probabilistic anonymity, (14) downgrading classifier effectiveness, (15) membership inference analysis, and (16) accurate predictions. Wagner *et al.* [124] summarized eighty privacy evaluation metrics used by different PPDP algorithms. The authors categorized the privacy metrics based on four common characteristics which are, (i) adversary models, (ii) data sources (i.e., auxiliary information), (iii) input for computation of metrics, and (iv) output measures. The selection of the metric depends on the data type, objectives of data publishing, and target application/users. However, in most cases, a single metric cannot capture the entire concept of privacy, therefore, two or more metrics are jointly used to measure the level of privacy offered by an anonymization algorithm/model. Recently, Zhao *et al.* [125] analyzed and discussed twenty-six different metrics used for privacy evaluation in anonymized graph. The authors suggested that no single metric is effective to evaluate privacy protection in anonymous graphs. Therefore, the authors suggest that the strengths of multiple privacy evaluation metrics can be combined to improve the overall measurement of user's privacy in a G'. This work employs multi-criteria decision analysis to the privacy measurement in a G', and it opens up a new research direction that may lead to significant improvements in future for privacy measurement.

B. DE-ANONYMIZATION METHODS EMPLOYED BY THE ADVERSARIES TO JEOPARDIZE USERS PRIVACY IN SN PUBLISHED DATA

Due to the phenomenal growth in SNs adoption around the globe, the SN data has become more reliable for conducting research and achieving multiple business/scientific objectives. The data-miners and analysts can extract the enormous amount of information embedded in the published G'. Aside from collecting the relevant information regarding some custom rules or desired communities from the published graph, the data-miners try to reveal true identities/private-information of the users for fulfilling multiple hidden objectives such as personalized service recommendation, user's profiling, and preference based digital contents selling. Interestingly, in some cases, the embedded knowledge extraction enables firms to be competitive in the market for long-run. Due to increase in information surges, availability of the various SNs, maturity of the machine learning and data mining tools, advancement in computing technologies, and attacker capabilities have made the personal information retrieval much easier. Due to the public access of the SNs and lack of user's awareness about the online privacy, the protection of privacy on the SN sites is very challenging. Therefore, it has become an active area of research in recent years. Adversaries are not only able to get multiple users information but also they can re-identify people uniquely with the help of multiple auxiliary sources. There exist several de-anonymization approaches in literature that were employed on the published graphs to infer the true identity or SA of the SN users. For instance, Azizy *et al.* [169] summarized and classified various de-anonymization approaches used by the adversaries. We summarize the recent de-anonymization approaches with examples in Figure 15.

Aside from the de-anonymization approaches explained above, Beigi *et al.* [170] summarized various seed-based and seed-free de-anonymization methods employed by the adversaries for privacy breaches in published SN data. Authors provided a detailed overview of latest SN data anonymization and de-anonymization approaches, and theoretical analysis about the graph's de-anonymization. In addition, various user's attributes inference methods have been reported by the authors with examples. Apart from the de-anonymization approaches explained in Figure 15 and previous studies,



Sr. #	Description of the de-anonymization approaches used by adversaries	
	Category	Description
1	Approach (Features used)	Graph Matching (network structure)
	Brief description of the approach (Examples)	It focuses on two graphs from the same SN service provider (SNSP) or perhaps from two different SNSP. Both graphs are used for mapping users between the shared anonymised nodes and measuring overlaps (knowledge graphs, sub-graphs, embedded graphs).
2	Approach (Features used)	Seed and grow (Growing links)
	Brief description of the approach (Examples)	It starts with a complex process called seeding to plant a node such as a user account in a SN graph, and then make it building up links with other nodes based on similarities (Profile attributes).
3	Approach (Features used)	Similarity matching (mapped graph features)
	Brief description of the approach(Examples)	It computes similar features between the target dataset/graphs and auxiliary information to perform the correct matching (user's attributes, and contents).
4	Approach (Features used)	Statistical matching (statistics of unique features)
	Brief description of the approach(Examples)	It depend on using known and acquired statistics from published graph. The attack relies on the unique features of users' data (i.e., friends) for re-identification (number of friends, virtual communities, online groups affiliation).
5	Approach (Features used)	Threading (difference in amount of information)
	Brief description of the approach(Examples)	It focuses on correlating the sequential releases of graphs and matching these releases with each other to identify newly-added/removed users (membership analysis).
6	Approach (Features used)	Link-based classifier (friendship, group membership)
	Brief description of the approach (Examples)	It analyses links and users groups information to infer friendship and group membership to identify some private attributes/users (social connections).
7	Approach (Features used)	Graph-based classifiers (group membership)
	Brief description of the approach (Examples)	It incorporates community structure information to predict the most probable classes/users for unlabeled/labeled nodes to re identify users uniquely (node identifiers, profiles).
8	Approach (Features used)	Similarity classifiers (activities, behaviours)
	Brief description of the approach (Examples)	It works on computing similarities of local features such as temporal activity, text, geographic, and social features to re-identify people across SNs (SN usage pattern, SN type, display name)
9	Approach (Features used)	Trials (visited locations, buying patterns)
	Brief description of the approach (Examples)	It works on brute force pattern to reveal user's identity and private information from G' (SN type, activities, profile name).
10	Approach (Features used)	Sparsity-based(unique value of attributes or dissimilar nodes)
	Brief description of the approach (Examples)	It works on finding rare values to correlated with some individuals to breach their privacy (less number of friends, location, unique race/religion).

FIGURE 15. Description about de-anonymization approaches used by the malevolent adversaries for privacy breaches.

we summarize the key items that enable unique identifications of an individual/groups or SA disclosures from the privacy preserved graph data publishing in Table 5.

These items are usually exploited by the adversaries to jeopardize user's privacy in SN published data to infer/predict true identities of users or their SAs. Each item assists in compromising an individual or community privacy when exploited by the adversaries. Various approaches based on these items/features have been proposed to compromise people's privacy by leveraging the single or multiple SNs' data (a.k.a across SNs). Drawing on the reviewed graph de-anonymization techniques, majority of the techniques enable adversaries to compromise users' privacy successfully.

The privacy items and corresponding de-anonymization methods summarized in Table 5 pose serious threats to the SN user's privacy in the PPGP. Apart from the well-known methods summarized above, the user's privacy can also be breached through the information acquisition about the changing interest of a user overtime and predictions about the user's private information through side channels of various types (i.e., interests). Recently, due to availability of the excessive amount of auxiliary information and advanced data mining tools, the scale and the scope of privacy breaches is expanding from an individual identification or SA disclosure to groups identity theft for achieving multiple scientific and business objectives, and identifying communities in SNs having common characteristics or common interests for accurate recommendations.

Furthermore, when a group of SN users form an online community, the SN service providers have access to more information including political viewpoints, preferences, relationship status or financial status because a user often readily shares more about him or herself in an online community. Accordingly, this goldmine of data when shared with the data-miners can lead to privacy breaches. In addition, social connection information prediction about a user with advanced data mining tools, and identifying an individual and users groups by contents analysis and activities has become serious challenge for SN service providers [122]. Hence, SNs users data sharing can endanger user's privacy in unexpected ways, and researchers are devising many domain and attacks specific, and general PPGP methods to combat with this uprising social problem (e.g., safeguarding SN users privacy).

TABLE 5. Description about the key items that are exploited by the adversaries to breach user's privacy.

Sr. No.	Items	Methods employed by the adversaries to breach user's privacy (a.k.a de-anonymization)	Representative Methods
1.	profile attributes	attributes coupling, computing attribute's similarities between published and auxiliary graphs, relevance analysis, quantifying the significance of attributes to invade user's privacy.	Yin et al. [171], Li et al. [172], Mao et al. [173], Zhang et al. [174, 175]
2.	user-generated contents in SN	measuring similarities of contents in terms of space, time and content dimensions, interest mapping, word embedding technologies on users messages.	Yongjun et al. [176], Nie et al. [177], Sha et al. [178]
3.	accounts on multiple social networks	pairwise identical factor graph models, identity search and matching by exploiting user-generated posts and number of friends.	Wang et al. [179], Ahmad et al. [180]
4.	online groups joined by a user on SNs	exploiting the group membership information of a user that he/she subscribe/join for information seeking or sharing and is available on the SN sites.	Wondracek et al. [181]
5.	communities a user is part of in the SNs	partitioning the whole networks/graphs into 'communities' and performs a mapping, exploiting the multi-hop neighborhood information with classifiers to breach privacy.	Nilizadeh et al. [182], Lee et al. [183]
6.	friends, neighbors' information	structure-based weighted neighborhood matching (SWNM), information shared by users' SN friends, friendship learning-based identification (FBI).	Fang et al. [184], Labitzke et al. [185], Youyang et al. [186]
7.	spatial and temporal information of the SN use	fusing temporal and spatial data with other related information such as check-in/posts data to infer private information of users.	Gao et al. [187]
8.	user's activities on the SNs sites	Combining multiple personal activities of a user such as posting content on SNs, commenting on SN sites, liking posts, and using mobile services.	Yoshiura et al. [188]
9.	display name of a user across SNs	Exploiting multiple characteristics of display names such as length similarity, letter distribution similarity and character similarity.	Li et al. [189]
10.	user's structural and content information	modeling the topics of user's interests from contents, capturing the interest-based and friend-based user co-occurrence in a G' to compromise user's privacy.	Wang et al. [190]
11.	tagging behavior of a user on SNs sites	extracting relevant features from the inconsistent tagging behaviors of a user and combining them with the profile attributes.	Zhao et al. [191]
12.	rating given by a user to a particular movie	extracting relevant features about the movies' rating, and gender estimation using traditional supervised classification methods.	Weinsberg et al. [192], Kosinski et al. [193]
13.	music choices	extraction of user's interests, semantic similarity computation and topic modeling based users' attribute prediction.	Chaabane et al. [194]
14.	social links (i.e., friends) and user behavior information	construction of the social behavior attribute network (SBA), integration of the related information in a unified framework, and inferring the target user's profile information via vote distribution attack (VIAL) model.	Gong et al. [195], [196]
15.	purchasing behavior	structured neural embedding (SNE) model to learn the representations from users' purchase data automatically to predict multiple demographic attributes. Five attributes (e.g., marital status, gender, income, education level, and age) of users were accurately predicted.	Wang et al. [197]
16.	networks of social attributes	modeling SN data as a social-attribute network (SAN) and exploiting the structural characteristics of the SAN, and computing attributes and nodes similarities to infer/predict the private links and user's attributes.	Jiang et al. [198], Thangam et al. [199], Song et al. [200]
17.	capturing user's activities/interactions	structural pattern learning, multi view modeling of both the target network and the auxiliary network, and employing target network reconstruction for inferring anonymized links.	Xian et al. [201]
18.	Exploiting vulnerable friend's information	finding and targeting the vulnerable friends who have applied insufficient privacy settings to infer their private information, and entire network of friends subsequently.	Gundecha et al. [202]
19.	Picture's metadata	mapping picture's meta-data to a low-dimensional vector space, and inferring user's gender subsequently.	Pijani et al. [203]
20.	utilizing both graph structure and user profile info.	reduce the size of candidate set by analyzing graph structure, and exploits user's profile information to identify the correct mapping of users with a very high confidence.	Li et al. [204]

C. STRUCTURAL ANONYMIZATION APPROACHES USED FOR APPLICATION-SPECIFIC SCENARIOS IN SN DATA PUBLISHING

In Table 4, we summarized the generic structural anonymization approaches used in the PPGP. In this subsection, we summarize the recent structural anonymization approaches used in application-specific scenarios of SN data publishing. These scenarios include, anonymization approaches for friends recommendation, community clustering/detection, collaborative filtering, topic modelling, and SN's users behavior analysis etc. Aside from the structural modifications in graphs, the anonymization approaches used in application-specific scenarios also consider the features of the applications for which the data is being anonymized. For example, Xu *et al.* [29] proposed a framework for discovering the privacy-preserved communities in the SNs. The proposed framework enables the formation/detection of different communities without revealing sensitive link information. In this work, we summarize the state-of-the art structural anonymization approaches used in application-specific scenarios of the SN in Table 6. The anonymization approaches summarized in Table 6 consider the features of the related application during anonymization process. Furthermore, some approaches have been used for multiple/heterogeneous applications due to overlapped properties/characteristics between them.

Recently, due to the phenomenal growth in opportunities offered by SNs data, researches have turned their attention to devise new and realistic anonymization methods leveraging advanced artificial intelligence (AI) techniques. Li *et al.* [226] designed a deep learning (DL) model that combines multiple factors such as attribute information, graph structure, and behaviour characteristics while avoiding tedious calculation procedures to measure the privacy in SNs. The proposed model considers the deep relationship between all three factors (user attributes information, graph structure, and behaviour characteristics) to accurately obtain the privacy score.

Alemany *et al.* [227] devised two metrics (Audience and Reachability) for assessment of privacy in information sharing scenario in the SNs. The authors performed rigorous simulations in different SN topologies and considering different layers, and concluded that network topology in SNs has a direct effect on the outreach of the information. Pensa *et al.* [228] described a knowledge-driven approach for enhancing privacy awareness in SNs. The proposed approach has ability to measure the privacy risk of the users and inform them whenever their privacy is breached or at risk. Also,

TABLE 6. Description about the anonymization approaches used in SN application-specific scenarios.

References	Proposed solutions	Main methods/concepts employed in the anonymization	Practical application scenario(s)
Xu et al. [29]	PPGP Framework	It utilizes both social connections and users' published contents during anonymization	Privacy preserved communities detection/discovery
Guo et al. [205]	Privacy preserving scheme	It utilizes SN users' social attributes and trust relationship during anonymization	Friend recommendation
Kukkala et al. [206]	Multiparty computation (MPC) protocols	It focuses on identifying the influential spreaders in the SN by securely performing k-Shell decomposition, PageRank and VoteRank algorithms.	Information diffusion
Dong et al. [207]	Matchmaking system	It encrypts users' identities, locations, as well as profiles to preserve the user's privacy in SNs	Finding potential friends/romantic partner
Zhang et al. [208]	Utility-based popularity anonymization scheme	It combines k-anonymous popularity-based following (KPF) protocol and fully utility-based interaction (FUI) protocols to protect users privacy.	targeted advertisements, opinion leader selection, information spreader
Georgiou et al. [209]	Algorithmic methodology	It alters an existing/determined community-aware trending topic structure so that it can preserve the privacy of the involved users while still reporting/modeling topics with a satisfactory level of utility	Trending topic detection in SNs.
Chinnaiah et al. [210]	Sanitization framework	It automatically preserves the user's privacy by detecting sensitive topic and minimizing the risk of sensitive information disclosure	Sensitive topic diffusion
Jiu-Ru et al. [211]	Novel framework	It uses k-automorphism model to protect the structural privacy and a cost-model based label generalization method to protect label privacy.	Subgraph pattern matching
Casas-Roma et al. [212]	PPDP framework	It uses a generic information loss measures (GIL), such as average distance or diameter, to evaluate to what extent the analysis of $G^{'}$ differs from the G to effectively preserve privacy.	Graph mining tasks
Dongsheng et al. [213]	Privacy-preserving recommender system	It organizes users into groups with diverse interests and users interact with the server via interest-specific pseudo users.	Privacy-preserving content recommendation
Mosallanezhad et al. [214]	Text annonymizer approach	It extracts a latent representation of the original data/text and use deep reinforcement learning for privacy and utility problem solving.	User behavioral modeling tasks
Tianchong et al. [215]	Anonymization algorithm	It adds and deletes enough edges to satisfy the privacy demand and use novel bottom-(ℓ, k) sketching to preserve user's privacy.	Analyzing the information transmission speed and rumor spreading models
Yiwei et al. [216]	Residual entropy minimization algorithm	It quantifies the amount of information revealed by a community structure and then solve this problem by residual entropy minimization (REM) algorithm to obfuscate a given community structure in a SN data.	Community structure deception
Ramezanian et al. [217]	AI-based privacy preserving system	It marks all subscribers as normal at the start, detect the bully message using artificial intelligence (AI) techniques, and protect victims automatically.	Cyberbullying prevention
Kavianpour et al. [218]	Privacy-preserving model	It provides a safe platform with high accuracy to detect malicious social interactions to enhance users' privacy in interactions with third-parties applications. It alleviates the possibility of information leakage and user's re-identification problems.	Controlling undesirable users' data diffusion to unauthorized third-parties
Rathore et al. [219]	Access control framework AppMonitor	It uses the relation-based access control (ReBAC) policy model based on the predicate calculus for regulating access of the users data with third-parties.	Restricting user's private data not to move outside of the SN platform
Boshrooyeh et al. [220]	Privado mechanism	It protects user privacy against active malicious adversaries on the SN sites using two non-colluding servers and honest but curious (HbC) adversarial model.	Users group-based advertising on SNs sites
Huang et al. [221]	platform-centric two-layer three-party game model	It protects user's privacy by analyzing the interactions among the three parties (user, platform, and adversary) by using a platform-centric two-layer three party game model.	Context-aware services on SNs sites, such as Facebook
Guo et al. [222]	attribute-based reputation value estimation system	It computes the reputation value of the users' attributes while preserving the privacy of the verifiability of the attributes and authenticity of the reputation value in SNs.	Content-based user's reputation estimation in SNs
Yang et al. [223]	DP Input and Manner (*DPI, DPM*) methods	It calculates the predicted scores and adopts two methods (*DPI, DPM*) of adding Laplace noise to hide individual ratings, and provide valuable prediction results to information consumers.	Collaborative filtering in recommender systems for social services (i.e., using SN data)
Aljably et al. [224]	privacy-preserving model based on local DP (LDP)	It protects user privacy in SNs by employing Laplace's probability distribution function (PDF) to generate random noise. Tt protects both profiles and activities of the users.	Anomaly detection in online social networks with high detection accuracy
Li et al. [225]	PPGP protocols based on homomorphic encryption	It prevents users information leakage during the graphs computing process (a.k.a privacy-aware graph computation) using homomorphic encryption technique.	social network analysis (e.g., graph operations) in multi-party computation process/setting

it helps the exposed users to customize their privacy level semi-automatically by limiting the number of manual operations. Li *et al.* [229] suggested that private information in the SNs sites is time-sensitive, which means that information held by the users who are no longer in the same environment/place as the target user may no longer be reliable/true and have lost its value. The authors combined behavioral characteristics and structural similarity to accurately filter the user groups who hold the current private information of a target user for measuring a user's privacy status. Ruggero *et al.* [230] suggested that user's privacy in SNs can be influenced by many external factors (e.g., the position of the user within the social graph, the relative risk of the network). To solve this problem, the authors devised a network-aware privacy score metric that accurately measure the user's privacy risk according to the characteristics of the network (i.e., G).

A semi-supervised framework based on structural embedding for account correlation was proposed by Zhou *et al.* [231]. It learns the latent and structural semantics for accounts correlation between networks. It correlates accounts with high accuracy by leveraging the semantic information among accounts through random walks approach. Furthermore, the user's identity disclosure/matches problems with limited profile items have also been reported in the literature [232], [233]. In recent years, federated learning (FL) based privacy preserving approaches have been proposed for anonymous data publishing with legitimate third-parties [234]–[238]. Furthermore, many anonymization approaches have been proposed to preserve the users' privacy in sequential publication of the SNs data [239]–[241].

D. CHALLENGES IN SOCIAL NETWORK DATA ANONYMIZATION

SN data is usually represented as a graph, and structural anonymization approach is applied to sanitize it before releasing with data-miners. Anonymizing SN data is much more challenging compared to tabular data due to complex structure and variety of information embedded in graphs about the entities (i.e., users). The three well-known challenges related to SN data anonymization are, (i) modeling background knowledge (BK) of the adversaries (in SNs, the appropriate modeling of the adversaries' BK is harder compared to the tabular data because adversaries can have access to the multiple pieces of information about an individual, and he/she can re-identify target individual from the G' by leveraging the BK.), (ii) devising a new structural anonymization method (devising a new anonymization method for SN data is very hard compared to the tabular data due to the structural dependence of entities on each other. In SN, a slight modification in the graph structure can affect the whole network. Hence, the adhoc solutions based on "divide and conquer" approach cannot be directly applied on the SN data), and (iii) quantifying the utility of the anonymous graph (in SN data, measuring the usefulness offered by an anonymous graph is not straightforward. The differences in G and G' properties are difficult to quantify. In addition, by adding new edges and vertices to increase privacy protection can often lead to excessive information loss in the PPGP.). Aside from the three key challenges explained above, the structural anonymization of massively large scale graphs involving many entities data is very challenging. Furthermore, in SN, amount and variety of data collected about entities increase exponentially with the passage of time. Thus, devising new structural anonymization approaches to solve these problems from both practical and theoretical perspectives have become imperative while benefiting from the SNs users data.

VI. SUMMARY AND DISCUSSION ABOUT THE PRIVACY ISSUES IN FUTURE COMPUTING PARADIGM

In this article, we have covered most of the concepts related to the anonymization approaches used for data owned by both physical organizations (e.g., hospitals, banks, and insurance companies etc.) and virtual platforms (e.g., Facebook, Twitter, and Link-din etc.). Specifically, we described the anonymization methods employed for two types of input data, tables and graphs. We emphasized more on the SN data anonymization considering the exponential adoption of the SNs around the globe by adults, and unprecedented opportunities these platforms offer in terms of business intelligence. Furthermore, the data-driven technologies and big data analytics are playing a vital role in extracting embedded knowledge from unstructured data to improve SQ [242].

Apart from these two types of data (e.g., tables and graphs), a wide range of data types such as matrix (e.g., trajectories information, market basket data, and ratings data), digital traces, logs, documents (e.g., medical prescriptions, disaster/disease control agencies data), images, videos, text documents (e.g., reviews, blogs, and opinions), and time series data can reveal private information in digital landscape. Hence, enterprises and organizations are constantly exploring new innovative strategies and methods to remain competitive in their market while ensuring users' privacy [243]–[247]. Nevertheless, data policies including privacy, intellectual property, security, and liability issues, should be addressed in all phases (e.g., collecting, pre-processing, anonymizing, sharing, and analytics) of person-specific data handling in order to exploit big data value. Decentralized anonymization methods are handy solutions to truly benefit from the data publishing without significantly impacting user's privacy [248], [249]. Nowadays, the AI techniques have become significantly mature to assist in data-driven decision making, users' privacy protection has become imperative for most organizations' success [250]. Considering the applications and unprecedented opportunities of data sharing, the privacy and utility trade-off resolution in PPDP remains challenging.

In future computing paradigm, four mainstream technologies will become the centre of the information technology (IT) world, big data, cloud computing, social networks, and internet of things (IoT). These technologies have ability to process any kind of data with advanced analytics tools to extract insights from collected data. The main drivers of these

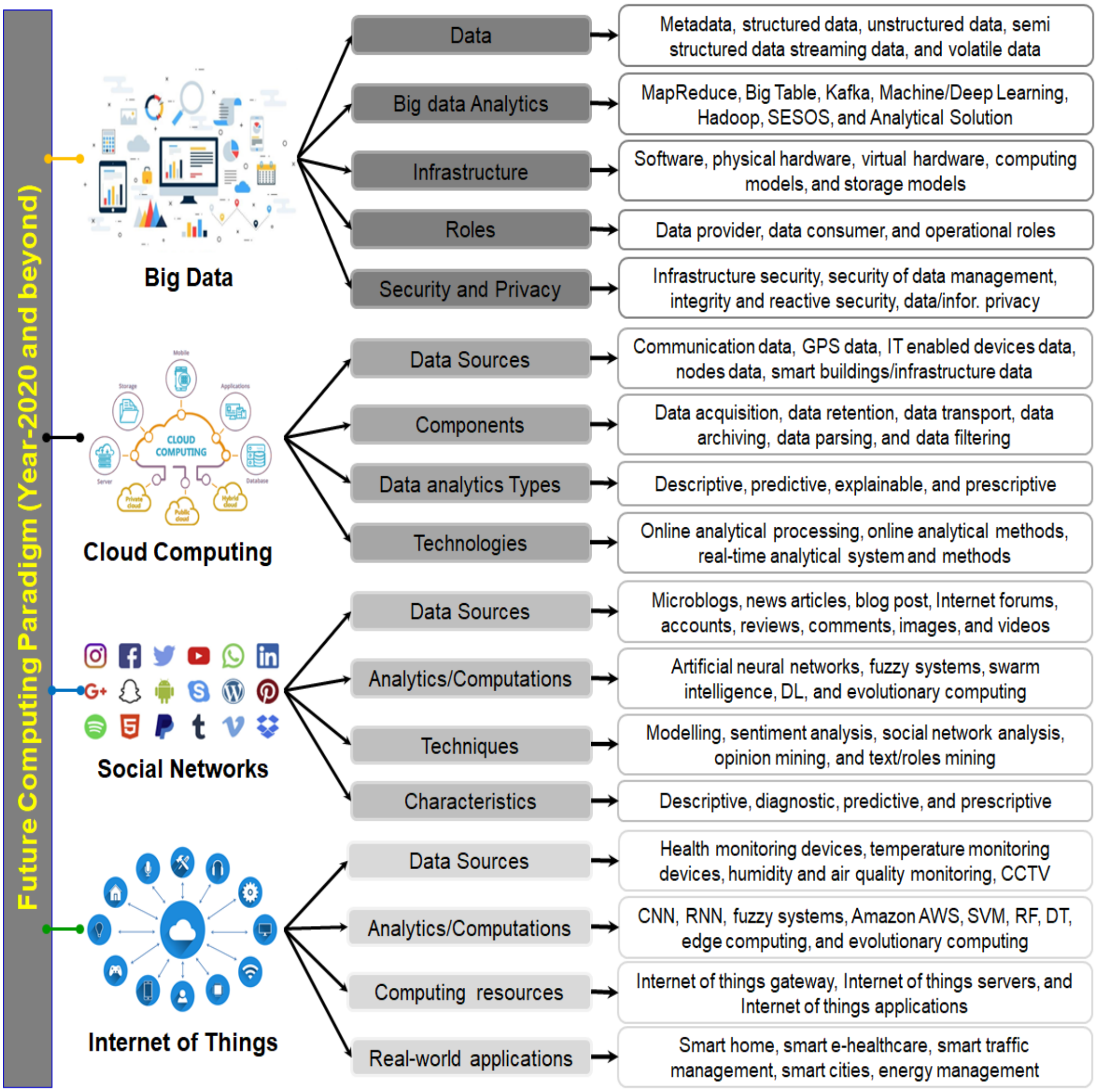

Future Computing Paradigm (Year-2020 and beyond)
Big Data
Data
Metadata, structured data, unstructured data, semi structured data streaming data, and volatile data
Big data Analytics
MapReduce, Big Table, Kafka, Machine/Deep Learning, Hadoop, SESOS, and Analytical Solution
Infrastructure
Software, physical hardware, virtual hardware, computing models, and storage models
Roles
Data provider, data consumer, and operational roles
Security and Privacy
Infrastructure security, security of data management, integrity and reactive security, data/infor. privacy
Cloud Computing
CLOUD COMPUTING
Data Sources
Communication data, GPS data, IT enabled devices data, nodes data, smart buildings/infrastructure data
Components
Data acquisition, data retention, data transport, data archiving, data parsing, and data filtering
Data analytics Types
Descriptive, predictive, explainable, and prescriptive
Technologies
Online analytical processing, online analytical methods, real-time analytical system and methods
Social Networks
Data Sources
Microblogs, news articles, blog post, Internet forums, accounts, reviews, comments, images, and videos
Analytics/Computations
Artificial neural networks, fuzzy systems, swarm intelligence, DL, and evolutionary computing
Techniques
Modelling, sentiment analysis, social network analysis, opinion mining, and text/roles mining
Characteristics
Descriptive, diagnostic, predictive, and prescriptive
Internet of Things
Data Sources
Health monitoring devices, temperature monitoring devices, humidity and air quality monitoring, CCTV
Analytics/Computations
CNN, RNN, fuzzy systems, Amazon AWS, SVM, RF, DT, edge computing, and evolutionary computing
Computing resources
Internet of things gateway, Internet of things servers, and Internet of things applications
Real-world applications
Smart home, smart e-healthcare, smart traffic management, smart cities, energy management

FIGURE 16. Detailed overview of the future computing paradigm and related concepts.

technologies are expanded internet connectivity, low cost sensors, higher mobile adoption, and large IoT investments. In Figure 16, we present the future computing paradigm's technologies, data sources, and analytics solutions that are in use to serve the mankind in better way compared to the recent past.

Despite the benefits offered by these latest technologies, there exist many barriers such as technological fragmentation, security concerns, implementation problems, and privacy concerns. Such potential barriers have been the bottleneck for the wide applications and development of these technologies and, thus, have attracted widespread concern. Among others, privacy concerns significantly hamper the development and applications of these technologies, and this field has attracted significant attention from the research community in recent years. Furthermore, many latest methodologies such as homomorphic encryption, federated learning, deep learning, and block chain have also been used for privacy protection in future computing paradigm (year 2020 and beyond) [251]–[258]. Butpheng *et al.* [259] presented various research perspectives related to privacy and security within IoT-cloud-based e-Health systems. Authors provided various benefits of IoT and cloud based e-Health systems, and analyzed security and privacy solutions in the study.

We summarize the various privacy issues related to the future computing paradigm after in-depth synthesis of the previous studies in Figure 17. We refer interested reader to previous works [254], [260]–[268] for more detailed descriptions and definitions of the privacy issues related to the future computing paradigm. Therefore, future research must be done to find new ways to make anonymization solutions more resilience towards these issues such as quantifying the risk and benefits of data publishing, deciding the appropriate mechanism for privacy preservation considering the adversaries' capabilities (e.g., worst case scenarios), verifying

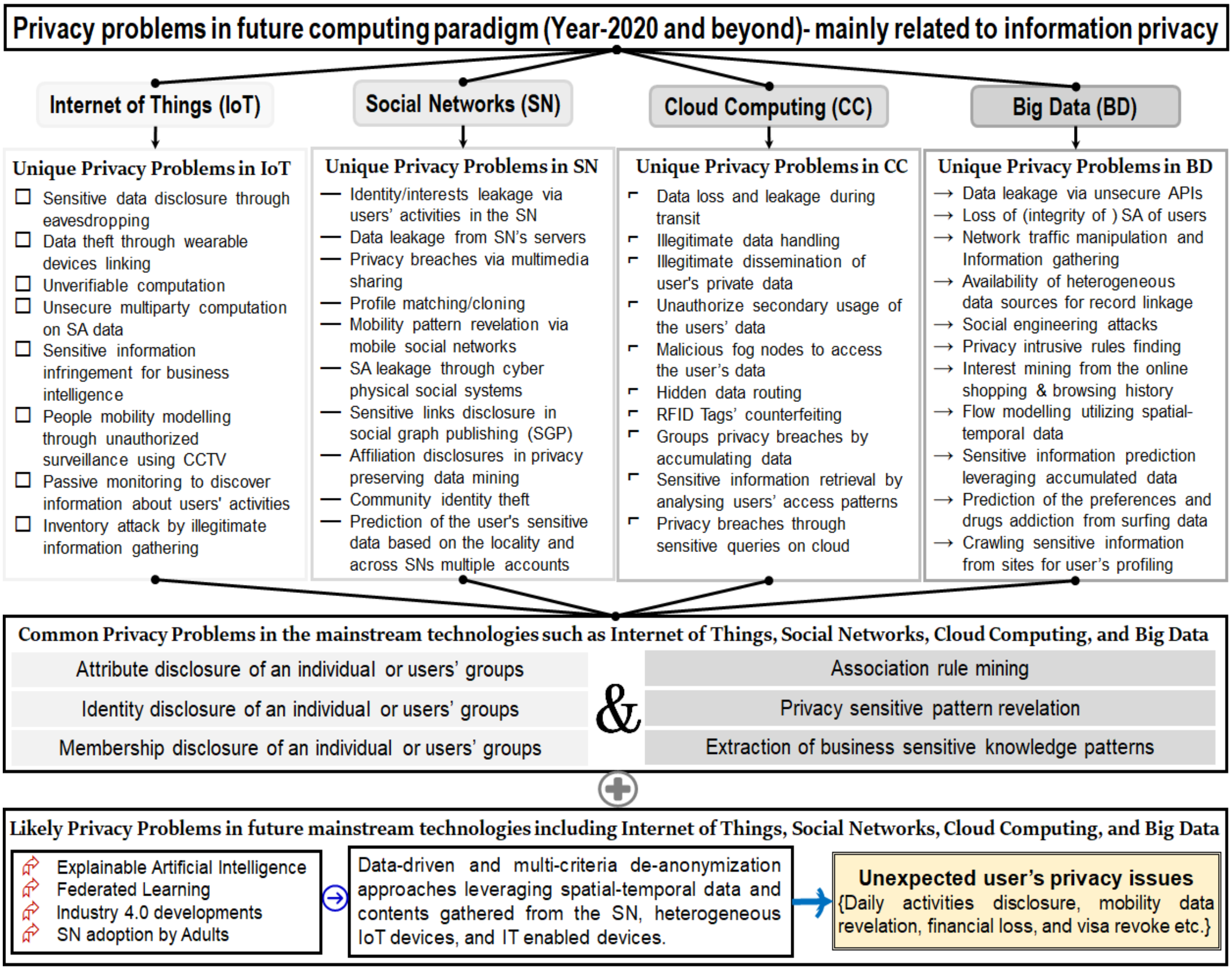

Privacy problems in future computing paradigm (Year-2020 and beyond)- mainly related to information privacy
Internet of Things (IoT)
Social Networks (SN)
Cloud Computing (CC)
Big Data (BD)
Unique Privacy Problems in IoT
☐ Sensitive data disclosure through eavesdropping
☐ Data theft through wearable devices linking
☐ Unverifiable computation
☐ Unsecure multiparty computation on SA data
☐ Sensitive information infringement for business intelligence
☐ People mobility modelling through unauthorized surveillance using CCTV
☐ Passive monitoring to discover information about users' activities
☐ Inventory attack by illegitimate information gathering
Unique Privacy Problems in SN
— Identity/interests leakage via users' activities in the SN
— Data leakage from SN's servers
— Privacy breaches via multimedia sharing
— Profile matching/cloning
— Mobility pattern revelation via mobile social networks
— SA leakage through cyber physical social systems
— Sensitive links disclosure in social graph publishing (SGP)
— Affiliation disclosures in privacy preserving data mining
— Community identity theft
— Prediction of the user's sensitive data based on the locality and across SNs multiple accounts
Unique Privacy Problems in CC
⌐ Data loss and leakage during transit
⌐ Illegitimate data handling
⌐ Illegitimate dissemination of user's private data
⌐ Unauthorize secondary usage of the users' data
⌐ Malicious fog nodes to access the user's data
⌐ Hidden data routing
⌐ RFID Tags' counterfeiting
⌐ Groups privacy breaches by accumulating data
⌐ Sensitive information retrieval by analysing users' access patterns
⌐ Privacy breaches through sensitive queries on cloud
Unique Privacy Problems in BD
→ Data leakage via unsecure APIs
→ Loss of (integrity of) SA of users
→ Network traffic manipulation and Information gathering
→ Availability of heterogeneous data sources for record linkage
→ Social engineering attacks
→ Privacy intrusive rules finding
→ Interest mining from the online shopping & browsing history
→ Flow modelling utilizing spatial-temporal data
→ Sensitive information prediction leveraging accumulated data
→ Prediction of the preferences and drugs addiction from surfing data
→ Crawling sensitive information from sites for user's profiling
Common Privacy Problems in the mainstream technologies such as Internet of Things, Social Networks, Cloud Computing, and Big Data
Attribute disclosure of an individual or users' groups
Identity disclosure of an individual or users' groups
Membership disclosure of an individual or users' groups
&
Association rule mining
Privacy sensitive pattern revelation
Extraction of business sensitive knowledge patterns
Likely Privacy Problems in future mainstream technologies including Internet of Things, Social Networks, Cloud Computing, and Big Data
Explainable Artificial Intelligence
Federated Learning
Industry 4.0 developments
SN adoption by Adults
Data-driven and multi-criteria de-anonymization approaches leveraging spatial-temporal data and contents gathered from the SN, heterogeneous IoT devices, and IT enabled devices.
Unexpected user's privacy issues
{Daily activities disclosure, mobility data revelation, financial loss, and visa revoke etc.}

FIGURE 17. Comprehensive overview of the privacy issues in the future computing paradigm (Year 2020 and beyond).

the effectiveness of privacy enhancing technologies through real-world applications or increasing users' awareness about the privacy leakage through hidden routes.

Moreover, achieving effective privacy protection should focus more on exploiting intrinsic characteristics of users' data or application features for which data is being anonymized by the simulation of diverse privacy approaches [269]–[271]. Furthermore, privacy enhancing technologies (PETs) for better privacy preservation in personalized services by satisfying both economical and ethical purposes have become more emergent than ever [272]. In recent years, privacy preserving machine learning (PPML) concept has been employed to extract the knowledge from distributed databases while ensuring data privacy [273], [274]. Due to the PPML, traditional machine learning algorithms can be adapted to secure users' data stored in multiple digital environments.

Recently, the privacy-aware data cleaning techniques have significantly reduced the data preparation cost of data analysis pipeline [275], [276]. These techniques allow the clients to buy clean, and curated data from heterogeneous service provider to perform analytics without compromising user's privacy. Furthermore, development of privacy information management system (PIMS) in accordance with international standard (e.g., ISO/IEC 27701) is imperative to safeguard the privacy of individuals or small groups in the population [277]. According to such system, if an anonymization mechanism cannot safeguard user's privacy or anonymized data can be used to identify individuals uniquely or small population groups, data cannot be released without legal advice or additional technical measures.

VII. PROMISING OPEN RESEARCH DIRECTIONS

Some promising open research directions/problems that need further research and developments from both academia and industry are outlined below.

- *Users groups' privacy issues*: In tabular data anonymization, majority of the existing approaches focus solely on an individual's privacy preservation. Thus, they are less resilient towards the users groups' privacy preservation.

For instance, k-anonymity model creates equivalence classes with k-users in each class. On the one hand, it protects individual privacy by hiding each user in other k-users' crowd. On the other hand, it explicitly discloses private information about users groups. Hence, devising new PPDP methods for users groups' privacy protection would be promising.

- *Excessive information loss caused by over-generalization of the QIs*: In tabular data anonymization, most of the existing anonymization approaches anonymize each QI present in a dataset that can lead to excessive information loss. As some QIs are not vulnerable in terms of user's privacy, and their unnecessary generalization significantly impact data utility. Thus, quantifying each QI impact on privacy and utility, and controlling unnecessary generalization to the extent possible while anonymizing user's data requires further research and developments from the research community.
- *Imbalanced datasets anonymization*: In some cases, the relational dataset can be highly imbalanced (i.e., the SA's values distribution is not uniform), and its anonymization is very challenging. In such datasets, enforcing hard constraints such as making each class ℓ-diverse or t-close is not possible in practice. Hence, it requires development of new approaches for anonymizing imbalanced datasets to effectively protect users' privacy without degrading data usefulness.
- *Effective resolution of privacy and utility trade-off in the PPDP*: In data anonymization, there exist a strong trade-off between privacy and utility. Tailoring the anonymization with privacy objectives can adversely affect the anonymous data utility, and vice-versa. This longstanding challenge in the field of PPDP seeking novel solutions to support privacy preserving big data analytics.
- *Personalized privacy preservation in SNs*: In SNs, each user has different requirements and concerns about his/her information privacy, which is called personalized privacy (PP). For example, in social graph some users may want to hide only sensitive relationship (i.e., lover) information, while some users may want to hide all social connections (i.e., all friends) information. Therefore, the PP involves high level of subjectiveness, and it is very difficult to implement. Hence, innovative solutions that can incorporate the SN users' PP requirements in SN data anonymization are required.
- *Accurate modeling of the adversaries' background knowledge*: Adversaries poses more side channel information as background knowledge (BK) about SNs users compared to the tabular data. This BK continues to grow due to the access to other publicly available SNs, and richness of information embedded in the SNs. Recently, text mining and natural language processing (NLP) techniques utilize the contents of SN users to match their identities. Thus, accurate modeling of the Bk while anonymizing SN data is very challenging, and further research on how to model the BK during anonymization process is required to effectively protect user's privacy in big data era.
- *Generic solutions for the social graph anonymization*: Generally, the SN data is modeled with the help of graphs (a.k.a. sociograms). These graphs can be of different types such as simple, directed, undirected, weighted, and labeled directed. The anonymization mechanism proposed for one type of the graph cannot be directly applied to the other. For instance, the k-degree anonymity concept cannot be applied to directed graphs straightforwardly as it requires the detailed analysis of in-and-out degree sequences. Hence, devising generic anonymization methods that can work with multiple graph's types will be an interesting research area in the future.
- *Controlling large scale user identification issues by evading data mining and SN analysis SNA) tools*: In SNs, users establish relationships with like-minded people or people having similar interests. This results into formation of the online communities. A growing body of research has been devoted to community discovery in the SNs. On the one hand, community discovery is beneficial for multiple purposes such as information spread and control. Moreover, community discovery in privacy preserved SN published data can jeopardize users and community privacy when published data is analyzed with advanced data mining and SNA tools. Hence, devising new solutions that are resilient towards community discovery and community-based node/user mapping in SN data publishing has become more pressing than ever.
- *Exploiting global and local features of SN data to safeguard against network reconciliation problems*: Recently, across SNs users identification by leveraging multiple methods such as multiple SNs graphs matching (i.e., network reconciliation), display names mapping, contents and activities analysis, accounts on the heterogeneous SNs, and their combinations has become an activate area of research. Accordingly, the privacy approaches need significant up-gradation to protect user's privacy in network reconciliation problems. In this regard, the approaches which perform data anonymization by exploiting local (i.e., common mapped neighbours, number of friends, and mutual friends etc.) and global (i.e., betweenness, centralities, ties strength, and multi-hop neighbour's information etc.) features of social graph will be an interesting research area in the near future for PPGP.
- *Metrics suites rather than single metric for quantifying the level of privacy in anonymous graphs*: Generally, one type of metric is employed for evaluating the level of privacy/utility in a G' offered by anonymization solutions. Moreover, in real world cases, the privacy quantified by one metric may not be monotonic (e.g., show lower

privacy results for stronger adversaries) or reliable from multiple viewpoints. Hence, combining multiple privacy metrics that can more accurately measure the level of privacy and can mitigate the weaknesses of individual metrics is a promising research direction in near future considering the widespread interest in SNs data publishing with legitimate information consumers.

- *Devising privacy-friendly mechanisms for exceptional situations*: During 2020, the whole world is facing an unanticipated and extraordinary challenge from an unknown enemy, called corona virus disease-19 (COVID-19) [278], [279]. The COVID-19 pandemic has affected every profession around the globe, and governments are heavily relying on the non-pharmaceutical interventions (e.g., strict lock-downs, cities and facilities closures, social distancing, geo-location based users' mobility analysis, proximity detection, and digital contact/suspect tracing etc.) to curb the spread of COVID-19 [280]. In addition, for epidemiological investigations, some governments employed extensive measures (e.g., credit card data, mobile phone signals, Bluetooth and GPS data, and CCTV data) to find the COVID-19's suspects and hotspots [281]–[283]. Due to the adoption of the digital methods, a huge amount of personal data has entered into the cyberspace, and privacy violations have been constantly reporting around the globe. For example, in Italy from January to April 2020, the privacy violations in healthcare sector related to companies and individuals have doubled [284]. Furthermore, the privacy violations in post COVID-19's era are expected to increase as many companies have collected the multi-fact data about the people lifestyles [285]–[287]. Considering the necessity of users' privacy preservation, the ethical aspects during and after COVID-19's era must be carefully observed and addressed [288]–[294]. Hence, privacy-friendly solutions are required for exceptional situations such as COVID-19 pandemic to safeguard patients' privacy and other pandemics' related aspects (e.g., privacy preserving symptoms tracking and reporting, collecting only relevant information from the users to protect privacy, encrypting sensitive information, GDPR-compliant and privacy-aware contact tracing, and decentralized solutions for computing the probability of exposure). Further, PPDP mechanism involving COVID-19's patients data to safeguard discrimination and hates towards certain religions, countries, sects, caste, and sexual minorities during published data analytics is deemed necessary. Hence, analyzing/solving the potential privacy risks and vulnerabilities in contact tracing apps developed by many countries to fight with the pandemic is a vibrant area of research.
- *Adoption of industry* 4.0 *techniques for the PPDP*: In recent years, many innovative techniques such as few-shot learning, federated learning, transfer learning, deep learning, and block-chain have revolutionized the human life in many aspects. These techniques have been extensively used in many areas such as healthcare, social engineering, data analytics, predictions and forecasting, knowledge extraction, and image recognition/analysis etc. Due to the availability of huge amount of labeled data, and ability to work in a decentralized fashion, these techniques can be utilized for users' privacy preservation with enhanced usefulness. The heterogeneous federated transfer learning (HFTL) framework [295], privacy-preserving deep learning (PPDL) technique [296], deep transfer learning (DTL) method [297], adaptive privacy preserving federated learning (APPFL) method [298], block-chain-enable privacy preserving (BCEPP) architectures [248], [299], secure collaborative few-shot learning (SCFSL) framework [300], searchable encryption (SE) methods leveraging ciphertext-policy attribute-based encryption (CP-ABE) [301], [302], data resource protection solution leveraging smart contracts [303], improving cyber security solutions utilizing AI's potential [304], and computational intelligence based methods for information security [305], to name a few have already been used in practical applications related to the PPDP. Hence, devising robust and lightweight techniques which involve less parameters and can co-work with the traditional anonymization approaches to scale up privacy preservation with enhanced data utility is a promising area of research for the future.

VIII. CONCLUSION

In this paper, we have presented the latest researches that have been proposed to release useful information while preserving user's privacy from malevolent adversaries, namely privacy preserving data publishing (PPDP). In recent years, there is an increasing focus on the rapid development of more practical anonymization solutions due to the significant rise in the privacy breaches across the globe, and this area is attracting researchers' interests drastically. Owing to the rapid technological developments in communication science and technology, tremendous amount of users' data can now be easily obtained in diverse formats, ranging from relational tables to complex social graphs. Although this increasing amount of data offers unprecedented opportunities for analytics, but it increases the chance of individuals' privacy breaches. In addition, most of the traditional anonymization algorithms that were proposed for the tabular data can rarely perform well on a social network (SN) (e.g., graphs) data without modifications. Hence, it is of paramount importance to provide good perspectives of the information privacy area involving both tabular and SN data along with recent anonymization researches. In this work, we have provided detailed and systemic coverage of the relational anonymization techniques used for the tabular data before presenting recent structural anonymization approaches used for SNs data anonymization. We have summarized and compared substantial number of anonymization approaches used for the information privacy protection involving both SNs and tabular data. Furthermore, we provide deeper insights on the privacy problems in future computing paradigm that will be helpful in devising more secure anonymization methods, and we discuss numerous promising open research directions/problems that need further research and developments. In this survey, we specifically focus on the SN data anonymization and de-anonymization techniques considering the widespread applications/use of the SNs data. In addition, SNs data contains a treasure of information that need to be protected from the malevolent adversaries. Thus, it indicates the ever increasing interests of researchers in the area of SN's data anonymization. Nevertheless, user's data anonymization is still irrefutably complex, and it requires significant improvements in existing approaches as well as devising new practical approaches with regard to better utility and privacy preservation. In future work, we are planning to devise new anonymization methods for SN data, and we intend to explore privacy problems in industry 4.0 technologies.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] K. Adhikari and R. K. Panda, "Users' information privacy concerns and privacy protection behaviors in social networks," *J. Global Marketing*, vol. 31, no. 2, pp. 96–110, Mar. 2018.

[2] J. Wieringa, P. K. Kannan, X. Ma, T. Reutterer, H. Risselada, and B. Skiera, "Data analytics in a privacy-concerned world," *J. Bus. Res.*, vol. 122, pp. 915–925, Jan. 2021.

[3] T. Li and N. Li, "On the tradeoff between privacy and utility in data publishing," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining - KDD*, 2009, pp. 517–526.

[4] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 838–852, Jun. 2013.

[5] B. C. Fung, K. Wang, A. W.-C. Fu, and S. Y. Philip, *Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques*. Boca Raton, FL, USA: CRC Press, 2010.

[6] R. K. Langari, S. Sardar, S. A. A. Mousavi, and R. Radfar, "Combined fuzzy clustering and firefly algorithm for privacy preserving in social networks," *Expert Syst. Appl.*, vol. 141, Art. no. 112968, 2020.

[7] X. Zhao, D. Pi, and J. Chen, "Novel trajectory privacy-preserving method based on clustering using differential privacy," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113241.

[8] J. Casas-Roma, "An evaluation of edge modification techniques for privacy-preserving on graphs," in *Proc. Int. Conf. Modeling Decisions Artif. Intell.*, Sant Cugat del Vallès, Spain: Springer, 2015, pp. 180–191.

[9] M. Rahimi, I. Sheikhbahaee UniversityIsfahan, M. Bateni, and H. Mohammadinejad, "Extended K-anonymity model for privacy preserving on micro data," *Int. J. Comput. Netw. Inf. Secur.*, vol. 7, no. 12, pp. 42–51, Nov. 2015.

[10] A. Anjum, K. K. R. Choo, A. Khan, A. Haroon, S. Khan, S. U. Khan, N. Ahmad, and B. Raza, "An efficient privacy mechanism for electronic health records," *Comput. Secur.*, vol. 72, pp. 196–211, Jan. 2018.

[11] N. Mohammed, R. Chen, B. C. M. Fung, and P. S. Yu, "Differentially private data release for data mining," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining - KDD*, 2011, pp. 493–501.

[12] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 10, no. 05, pp. 557–570, Oct. 2002.

[13] A. Friedman, R. Wolff, and A. Schuster, "Providing k-anonymity in data mining," *VLDB J.*, vol. 17, no. 4, pp. 789–804, Jul. 2008.

[14] B. C. M. Fung, K. Wang, and P. S. Yu, "Anonymizing classification data for privacy preservation," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 5, pp. 711–725, May 2007.

[15] S. Kisilevich, L. Rokach, Y. Elovici, and B. Shapira, "Efficient multidimensional suppression for K-Anonymity," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 3, pp. 334–347, Mar. 2010.

[16] J. Li, J. Liu, M. Baig, and R. C.-W. Wong, "Information based data anonymization for classification utility," *Data Knowl. Eng.*, vol. 70, no. 12, pp. 1030–1045, Dec. 2011.

[17] P. Geetha, C. Naikodi, and S. L. N. Setty, "Design of big data privacy framework—A balancing act," in *Advances in Data Sciences, Security and Applications*. Singapore: Springer, 2020, pp. 253–265.

[18] C. Dwork, "Differential privacy: A survey of results," in *Proc. Int. Conf. Theory Appl. Models Comput.* Changsha, China: Springer, 2008, pp. 1–19.

[19] A. Friedman and A. Schuster, "Data mining with differential privacy," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining - KDD*, 2010, pp. 493–502.

[20] J. Soria-Comas, J. Domingo-Ferrer, D. Sanchez, and D. Megias, "Individual differential privacy: A utility-preserving formulation of differential privacy guarantees," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 6, pp. 1418–1429, Jun. 2017.

[21] R. Sarathy and K. Muralidhar, "Evaluating laplace noise addition to satisfy differential privacy for numeric data," *Trans. Data Privacy*, vol. 4, no. 1, pp. 1–17, Apr. 2011

[22] J. Hua, A. Tang, Y. Fang, Z. Shen, and S. Zhong, "Privacy-preserving utility verification of the data published by non-interactive differentially private mechanisms," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2298–2311, Oct. 2016.

[23] U. Kumaran, "A secure and privacy-preserving approach to protect user data across cloud based online social networks," *Int. J. Grid High Perform. Comput.*, vol. 12, no. 2, pp. 1–24, 2020.

[24] X. Li, J. Yang, Z. Sun, and J. Zhang, "Differential privacy for edge weights in social networks," *Secur. Commun. Netw.*, vol. 2017, pp. 1–10, Mar. 2017.

[25] N. Yazdanjue, M. Fathian, and B. Amiri, "Evolutionary algorithms for k-Anonymity in social networks based on clustering approach," *Comput. J.*, vol. 63, no. 7, pp. 1039–1062, Jul. 2020.

[26] Y. Hao, H. Cao, C. Hu, K. Bhattarai, and S. Misra, "K-anonymity for social networks containing rich structural and textual information," *Social Netw. Anal. Mining*, vol. 4, no. 1, p. 223, Dec. 2014.

[27] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore art thou r3579x? anonymized social networks, hidden patterns, and structural steganography," in *Proc. 16th Int. Conf. World Wide Web*, 2007, pp. 181–190.

[28] J. Casas-Roma, J. Salas, F. D. Malliaros, and M. Vazirgiannis, "K-degree anonymity on directed networks," *Knowl. Inf. Syst.*, vol. 61, no. 3, pp. 1743–1768, Dec. 2019.

[29] X. Zheng, Z. Cai, G. Luo, L. Tian, and X. Bai, "Privacy-preserved community discovery in online social networks," *Future Gener. Comput. Syst.*, vol. 93, pp. 1002–1009, Apr. 2019.

[30] J. Casas-Roma, J. Herrera-Joancomartí, and V. Torra, "A survey of graph-modification techniques for privacy-preserving on networks," *Artif. Intell. Rev.*, vol. 47, no. 3, pp. 341–366, Mar. 2017.

[31] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey of recent developments," *ACM Comput. Surv.*, vol. 42, no. 4, pp. 1–53, Jun. 2010.

[32] K. Rajendran, M. Jayabalan, and M. E. Rana, "A study on k-anonymity, l-diversity, and t-closeness techniques," *IJCSNS*, vol. 17, no. 12, p. 172, 2017.

[33] A. Gkoulalas-Divanis, G. Loukides, and J. Sun, "Publishing data from electronic health records while preserving privacy: A survey of algorithms," *J. Biomed. Informat.*, vol. 50, pp. 4–19, Aug. 2014.

[34] H.-Y. Tran and J. Hu, "Privacy-preserving big data analytics a comprehensive survey," *J. Parallel Distrib. Comput.*, vol. 134, pp. 207–218, Dec. 2019.

[35] A. Sharma, G. Singh, and S. Rehman, "A review of big data challenges and preserving privacy in big data," in *Advances in Data and Information Sciences*. Singapore: Springer, 2020, pp. 57–65.

[36] M. Binjubeir, A. A. Ahmed, M. A. B. Ismail, A. S. Sadiq, and M. Khurram Khan, "Comprehensive survey on big data privacy protection," *IEEE Access*, vol. 8, pp. 20067–20079, 2020.

[37] Y. Mengmeng, Z. Tianqing, Z. Wanlei, and X. Yang, "Attacks and countermeasures in social network data publishing," *ZTE Commun.*, vol. 14, no. S0, pp. 2–9, 2019.

[38] M. Siddula, L. Li, and Y. Li, "An empirical study on the privacy preservation of online social networks," *IEEE Access*, vol. 6, pp. 19912–19922, 2018.

[39] B. Zhou, J. Pei, and W. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," *ACM SIGKDD Explor. Newslett.*, vol. 10, no. 2, pp. 12–22, Dec. 2008.

[40] E. Zheleva and L. Getoor, "Privacy in social networks: A survey," in *Social Network Data Analytics*. Boston, MA, USA: Springer, 2011, pp. 277–306.

[41] J. Abawajy and M. I. H. Ninggal, "Tutut herawan privacy preserving social network data publication," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1974–1997, 3rd Quart., 2016, doi: 10.1109/COMST.2016.2533668.

[42] L. Xing, K. Deng, H. Wu, P. Xie, H. V. Zhao, and F. Gao, "A survey of across social networks user identification," *IEEE Access*, vol. 7, pp. 137472–137488, 2019.

[43] A. Esfandyari, M. Zignani, S. Gaito, and G. P. Rossi, "User identification across online social networks in practice: Pitfalls and solutions," *J. Inf. Sci.*, vol. 44, no. 3, pp. 377–391, Jun. 2018.

[44] J. Casas-Roma, "An evaluation of vertex and edge modification techniques for privacy-preserving on graphs," *J. Ambient Intell. Humanized Comput.*, pp. 1–17, Jun. 2019, doi: 10.1007/s12652-019-01363-6.

[45] S. Badsha, X. Yi, and I. Khalil, "A practical privacy-preserving recommender system," *Data Sci. Eng.*, vol. 1, no. 3, pp. 161–177, Sep. 2016.

[46] R. Mendes and J. P. Vilela, "Privacy-preserving data mining: Methods, metrics, and applications," *IEEE Access*, vol. 5, pp. 10562–10582, 2017.

[47] M. Jayabalan and M. E. Rana, "Anonymizing healthcare records: A study of privacy preserving data publishing techniques," *Adv. Sci. Lett.*, vol. 24, no. 3, pp. 1694–1697, Mar. 2018.

[48] A. Vedangi and V. Anandam, "Data slicing technique to privacy preserving and data publishing," *Cancer*, vol. 4790, no. 4790, p. 4790, 2013.

[49] M. E. Nergiz, M. Atzori, and C. Clifton, "Hiding the presence of individuals from shared databases," in *Proc. ACM SIGMOD Int. Conf. Manage. Data SIGMOD*, 2007, pp. 665–676.

[50] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: Privacy and data mining," *IEEE Access*, vol. 2, pp. 1149–1176, 2014.

[51] R. Khan, X. Tao, A. Anjum, H. Sajjad, S. U. R. Malik, A. Khan, and F. Amiri, "Privacy preserving for multiple sensitive attributes against fingerprint correlation attack satisfying c-Diversity," *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 1–18, Jan. 2020.

[52] C. Watanabe, T. Amagasa, and L. Liu, "Privacy risks and countermeasures in publishing and mining social network data," in *Proc. 7th Int. Conf. Collaborative Comput., Netw., Appl. Worksharing*, 2011, pp. 55–66.

[53] M. Kiabod, M. N. Dehkordi, and B. Barekatain, "TSRAM: A time-saving k-degree anonymization method in social network," *Expert Syst. Appl.*, vol. 125, pp. 378–396, Jul. 2019.

[54] V. V. H. Pham, S. Yu, K. Sood, and L. Cui, "Privacy issues in social networks and analysis: A comprehensive survey," *IET Netw.*, vol. 7, no. 2, pp. 74–84, Mar. 2018.

[55] A. Maurya and M. Singh, "A survey on social networks: Issues and attacks," in *Recent Advances in Mathematics, Statistics and Computer Science*. Singapore: World Scientific, 2016, pp. 634–642.

[56] F. Li, H. Li, B. Niu, and J. Chen, "Privacy computing: Concept, computing framework, and future development trends," *Engineering*, vol. 5, no. 6, pp. 1179–1192, Dec. 2019.

[57] A. Zigomitros, F. Casino, A. Solanas, and C. Patsakis, "A survey on privacy properties for data publishing of relational data," *IEEE Access*, vol. 8, pp. 51071–51099, 2020.

[58] S. M. Shah and R. A. Khan, "Secondary use of electronic health record: Opportunities and challenges," 2020, *arXiv:2001.09479*. [Online]. Available: http://arxiv.org/abs/2001.09479

[59] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowl. Discovery From Data*, vol. 1, no. 1, p. 3, 2007.

[60] N. Li, T. Li, and S. Venkatasubramanian, "T-closeness: Privacy beyond k-Anonymity and l-Diversity," in *Proc. IEEE 23rd Int. Conf. Data Eng.*, Apr. 2007, pp. 106–115.

[61] J. S. Davis and O. Osoba, "Improving privacy preservation policy in the modern information age," *Health Technol.*, vol. 9, no. 1, pp. 65–75, Jan. 2019.

[62] X. Ding, W. Yang, K.-K. Raymond Choo, X. Wang, and H. Jin, "Privacy preserving similarity joins using MapReduce," *Inf. Sci.*, vol. 493, pp. 20–33, Aug. 2019.

[63] T. M. Truta and B. Vinay, "Privacy protection: P-Sensitive k-Anonymity property," in *Proc. 22nd Int. Conf. Data Eng. Workshops (ICDEW)*, Apr. 2006, p. 94.

[64] X. Sun, L. Sun, and H. Wang, "Extended k-anonymity models against sensitive attribute disclosure," *Comput. Commun.*, vol. 34, no. 4, pp. 526–535, Apr. 2011.

[65] R. C.-W. Wong, J. Li, A. W.-C. Fu, and K. Wang, "(α, k)-anonymity: An enhanced k-anonymity model for privacy preserving data publishing," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2006, pp. 754–759.

[66] H. Jian-min, Y. Hui-qun, Y. Juan, and C. Ting-ting, "A complete (alpha,k)-anonymity model for sensitive values individuation preservation," in *Proc. Int. Symp. Electron. Commerce Secur.*, Aug. 2008, pp. 318–323.

[67] T. Khanh Dang, J. Ku, and H. V. Q. Phuong, "Protecting privacy while discovering and maintaining association rules," in *Proc. 4th IFIP Int. Conf. New Technol., Mobility Secur.*, Feb. 2011, pp. 1–5.

[68] N. Maheshwarkar, K. Pathak, and N. S. Choudhari, "K-anonymity model for multiple sensitive attributes," *Int. J. Comput. Appl.*, vol. 1, no. 1, pp. 51–56, 2012.

[69] M. Nithya and T. Sheela, "Predictive delimiter for multiple sensitive attribute publishing," *Cluster Comput.*, vol. 22, no. S5, pp. 12297–12304, Sep. 2019.

[70] Widodo, E. K. Budiardjo, W. C. Wibowo, and H. T. Y. Achsan, "An approach for distributing sensitive values in k-Anonymity," in *Proc. Int. Workshop Big Data Inf. Secur. (IWBIS)*, Oct. 2019, pp. 109–114.

[71] Widodo, E. K. Budiardjo, and W. C. Wibowo, "Privacy preserving data publishing with multiple sensitive attributes based on overlapped slicing," *Information*, vol. 10, no. 12, p. 362, Nov. 2019.

[72] L. Zhang, J. Xuan, R. Si, and R. Wang, "An improved algorithm of individuation K-Anonymity for multiple sensitive attributes," *Wireless Pers. Commun.*, vol. 95, no. 3, pp. 2003–2020, Aug. 2017.

[73] C. Liu, S. Chen, S. Zhou, J. Guan, and Y. Ma, "A novel privacy preserving method for data publication," *Inf. Sci.*, vol. 501, pp. 421–435, Oct. 2019.

[74] Y.-C. Tsai, S.-L. Wang, I.-H. Ting, and T.-P. Hong, "Flexible sensitive k-anonymization on transactions," *World Wide Web*, vol. 23, pp. 1–16, Apr. 2020.

[75] H. Zhu, S. Tian, and K. Lu, "Privacy-preserving data publication with features of independent-diversity," *Comput. J.*, vol. 58, no. 4, pp. 549–571, Apr. 2015.

[76] E. Elabd, H. Abdulkader, and A. Mubark, "L–diversity-based semantic anonymaztion for data publishing," *Int. J. Inf. Technol. Comput. Sci.*, vol. 7, no. 10, pp. 1–7, Sep. 2015.

[77] L. Zhang, L. Wang, S. Jajodia, and A. Brodsky, "L-cover: Preserving diversity by anonymity," in *Workshop on Secure Data Management*. Trento, Italy: Springer, 2009, pp. 158–171.

[78] H. Tian and W. Zhang, "Extending ℓ-diversity to generalize sensitive data," *Data Knowl. Eng.*, vol. 70, no. 1, pp. 101–126, Jan. 2011.

[79] B. K. Tripathy, A. Maity, B. Ranajit, and D. Chowdhuri, "A fast p-sensitive l-diversity anonymisation algorithm," in *Proc. IEEE Recent Adv. Intell. Comput. Syst.*, Sep. 2011, pp. 741–744.

[80] S. Chakraborty and B. Tripathy, "Privacy preservation in relational data through l-diversity and recursive (c, l) diversity anonymisation," *Int. J. Math. Model. Numer. Optim.*, vol. 7, nos. 3–4, pp. 338–362, 2016.

[81] X. Wu, Y. Wei, T. Jiang, Y. Wang, and S. Jiang, "A micro-aggregation algorithm based on density partition method for anonymizing biomedical data," *Current Bioinf.*, vol. 14, no. 7, pp. 667–675, Sep. 2019.

[82] Y. Xiao and H. Li, "Privacy preserving data publishing for multiple sensitive attributes based on security level," *Information*, vol. 11, no. 3, p. 166, Mar. 2020.

[83] F. Ashkouti, K. Khamforoosh, and A. Sheikhahmadi, "DI-mondrian: Distributed improved Mondrian for satisfaction of the L-diversity privacy model using apache spark," *Inf. Sci.*, vol. 546, pp. 1–24, Feb. 2021.

[84] K. Oishi, Y. Sei, Y. Tahara, and A. Ohsuga, "Semantic diversity: Privacy considering distance between values of sensitive attribute," *Comput. Secur.*, vol. 94, Jul. 2020, Art. no. 101823.

[85] N. Li, T. Li, and S. Venkatasubramanian, "Closeness: A new privacy measure for data publishing," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 7, pp. 943–956, Jul. 2010.

[86] M. Wang, Z. Jiang, Y. Zhang, and H. Yang, "T-closeness slicing: A new privacy-preserving approach for transactional data publishing," *Informs J. Comput.*, vol. 30, no. 3, pp. 438–453, Aug. 2018.

[87] Y. Qu, J. Xu, and S. Yu, “Privacy preserving in big data sets through multiple shuffle,” in *Proc. Australas. Comput. Sci. Week Multiconf.*, Jan. 2017, pp. 1–8.

[88] R. Wang, Y. Zhu, T.-S. Chen, and C.-C. Chang, “Privacy-preserving algorithms for multiple sensitive attributes satisfying t-Closeness,” *J. Comput. Sci. Technol.*, vol. 33, no. 6, pp. 1231–1242, Nov. 2018.

[89] J. Cao, P. Karras, P. Kalnis, and K.-L. Tan, “SABRE: A sensitive attribute bucketization and REdistribution framework for t-closeness,” *VLDB J.*, vol. 20, no. 1, pp. 59–81, Feb. 2011.

[90] J. Soria-Comas, J. Domingo-Ferrer, D. Sanchez, and S. Martinez, “T-closeness through microaggregation: Strict privacy with enhanced utility preservation,” *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3098–3110, Nov. 2015.

[91] Y. Sei, H. Okumura, T. Takenouchi, and A. Ohsuga, “Anonymization of sensitive quasi-identifiers for l-Diversity and t-Closeness,” *IEEE Trans. Depend. Sec. Comput.*, vol. 16, no. 4, pp. 580–593, Jul. 2019.

[92] E. Weber, “A method for t-close anonymization in the presence of multiple numerical sensitive attributes,” Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Wichita State Univ., Wichita, Kansas, 2020.

[93] K. Nissim, S. Raskhodnikova, and A. Smith, “Smooth sensitivity and sampling in private data analysis,” in *Proc. 39th Annu. ACM Symp. Theory Comput. - STOC*, 2007, pp. 75–84.

[94] A. Korolova, K. Kenthapadi, N. Mishra, and A. Ntoulas, “Releasing search queries and clicks privately,” in *Proc. 18th Int. Conf. World Wide Web - WWW*, 2009, pp. 171–180.

[95] D. Kifer and A. Machanavajjhala, “No free lunch in data privacy,” in *Proc. Int. Conf. Manage. Data - SIGMOD*, 2011, pp. 193–204.

[96] L. Yu, L. Liu, C. Pu, M. E. Gursoy, and S. Truex, “Differentially private model publishing for deep learning,” in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2019, pp. 332–349.

[97] D. Lv and S. Zhu, “Achieving correlated differential privacy of big data publication,” *Comput. Secur.*, vol. 82, pp. 184–195, May 2019.

[98] M. Hardt, K. Ligett, and F. McSherry, “A simple and practical algorithm for differentially private data release,” in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 2339–2347.

[99] J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez, and S. Martínez, “Enhancing data utility in differential privacy via microaggregation-based kk -anonymity,” *VLDB J.*, vol. 23, no. 5, pp. 771–794, Oct. 2014.

[100] H. Lee and Y. D. Chung, “Differentially private release of medical microdata: An efficient and practical approach for preserving informative attribute values,” *BMC Med. Informat. Decis. Making*, vol. 20, no. 1, pp. 1–15, Dec. 2020.

[101] H. B. Kartal, X. Liu, and X.-B. Li, “Differential privacy for the vast majority,” *ACM Trans. Manage. Inf. Syst.*, vol. 10, no. 2, pp. 1–15, Aug. 2019.

[102] N. Wang and M. S. Kankanhalli, “Protecting sensitive place visits in privacy-preserving trajectory publishing,” *Comput. Secur.*, vol. 97, Oct. 2020, Art. no. 101949.

[103] F. Farokhi, “Privacy-preserving public release of datasets for support vector machine classification,” *IEEE Trans. Big Data*, early access, Jan. 3, 2020, doi: 10.1109/TBDATA.2019.2963391.

[104] A. Majeed, F. Ullah, and S. Lee, “Vulnerability- and diversity-aware anonymization of personally identifiable information for improving user privacy and utility of publishing data,” *Sensors*, vol. 17, no. 5, p. 1059, May 2017.

[105] J. Soria-Comas and J. Domingo-Ferrert, “Differential privacy via t-closeness in data publishing,” in *Proc. 11th Annu. Conf. Privacy, Secur. Trust*, Jul. 2013, pp. 27–35.

[106] M. R. Sarrafi Aghdam and N. Sonehara, “Achieving high data utility K-Anonymization using similarity-based clustering model,” *IEICE Trans. Inf. Syst.*, vol. E99.D, no. 8, pp. 2069–2078, 2016.

[107] A. Majeed and S. Lee, “Attribute susceptibility and entropy based data anonymization to improve users community privacy and utility in publishing data,” *Appl. Intell.*, vol. 50, pp. 1–20, Mar. 2020.

[108] S. Kabou, S. M. Benslimane, and M. Mosteghanemi, “A survey on privacy preserving dynamic data publishing,” *Int. J. Organizational Collective Intell.*, vol. 8, no. 4, pp. 1–20, Oct. 2018.

[109] Y. Shi, Z. Zhang, H.-C. Chao, and B. Shen, “Data privacy protection based on micro aggregation with dynamic sensitive attribute updating,” *Sensors*, vol. 18, no. 7, p. 2307, Jul. 2018.

[110] X. Xiao and Y. Tao, “M-invariance: Towards privacy preserving republication of dynamic datasets,” in *Proc. ACM SIGMOD Int. Conf. Manage. Data - SIGMOD*, 2007, pp. 689–700.

[111] A. Anjum, G. Raschia, M. Gelgon, A. Khan, S. U. R. Malik, N. Ahmad, M. Ahmed, S. Suhail, and M. M. Alam, “τ-safety: A privacy model for sequential publication with arbitrary updates,” *Comput. Secur.*, vol. 66, pp. 20–39, May 2017.

[112] H. Zhu, H.-B. Liang, L. Zhao, D.-Y. Peng, and L. Xiong, “τ-safe (l, k)-Diversity privacy model for sequential publication with high utility,” *IEEE Access*, vol. 7, pp. 687–701, 2019.

[113] M. E. Nergiz and C. Clifton, “““δ-presence without complete world knowledge,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 6, pp. 868–883, Jun. 2010.

[114] B. Ouafae, R. Mariam, L. Oumaima, and L. Abdelouahid, “Data anonymization in social networks state of the art, exposure of shortcomings and discussion of new innovations,” in *Proc. 1st Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET)*, Apr. 2020, pp. 1–10.

[115] S. Kumar and P. Kumar, “Upper approximation based privacy preserving in online social networks,” *Expert Syst. Appl.*, vol. 88, pp. 276–289, Dec. 2017.

[116] J. Vadisala and V. K. Vatsavayi, “Challenges in social network data privacy,” *Int. J. Comput. Intell. Res.*, vol. 13, no. 5, pp. 965–979, 2017.

[117] T. Tassa and D. J. Cohen, “Anonymization of centralized and distributed social networks by sequential clustering,” *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 2, pp. 311–324, Feb. 2013.

[118] X. Ding, X. Zhang, Z. Bao, and H. Jin, “Privacy-preserving triangle counting in large graphs,” in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 1283–1292.

[119] A. J. O’Malley and J.-P. Onnela, *Introduction to Social Network Analysis*. New York, NY: Springer, 2019, pp. 617–660.

[120] R. Pinot, “Minimum spanning tree release under differential privacy constraints,” 2018, *arXiv:1801.06423*. [Online]. Available: http://arxiv.org/abs/1801.06423

[121] S. Bourahla, M. Laurent, and Y. Challal, “Privacy preservation for social networks sequential publishing,” *Comput. Netw.*, vol. 170, Apr. 2020, Art. no. 107106.

[122] K. Deng, L. Xing, L. Zheng, H. Wu, P. Xie, and F. Gao, “A user identification algorithm based on user behavior analysis in social networks,” *IEEE Access*, vol. 7, pp. 47114–47123, 2019.

[123] S. Ji, P. Mittal, and R. Beyah, “Graph data anonymization, de-anonymization attacks, and de-anonymizability quantification: A survey,” *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1305–1326, 2nd Quart., 2017.

[124] I. Wagner and D. Eckhoff, “Technical privacy metrics: A systematic survey,” *ACM Comput. Surv.*, vol. 51, no. 3, pp. 1–38, Jul. 2018.

[125] Y. Zhao and I. Wagner, “Using metrics suites to improve the measurement of privacy in graphs,” *IEEE Trans. Depend. Sec. Comput.*, early access, Mar. 13, 2020, doi: 10.1109/TDSC.2020.2980271.

[126] T. Ma, Y. Zhang, J. Cao, J. Shen, M. Tang, Y. Tian, A. Al-Dhelaan, and M. Al-Rodhaan, “KDVEM: A k-degree anonymity with vertex and edge modification algorithm,” *Computing*, vol. 97, no. 12, pp. 1165–1184, 2015.

[127] J. Casas-Roma, J. Herrera-Joancomartí, and V. Torra, “K-degree anonymity and edge selection: Improving data utility in large networks,” *Knowl. Inf. Syst.*, vol. 50, no. 2, pp. 447–474, Feb. 2017.

[128] Y. Wang and B. Zheng, “Preserving privacy in social networks against connection fingerprint attacks,” in *Proc. IEEE 31st Int. Conf. Data Eng.*, Apr. 2015, pp. 54–65.

[129] M. I. H. Ninggal and J. H. Abawajy, “Utility-aware social network graph anonymization,” *J. Netw. Comput. Appl.*, vol. 56, pp. 137–148, Oct. 2015.

[130] A. Milani Fard and K. Wang, “Neighborhood randomization for link privacy in social network analysis,” *World Wide Web*, vol. 18, no. 1, pp. 9–32, Jan. 2015.

[131] P. Liu, L.-E. Wang, and X. Li, “Randomized perturbation for privacy-preserving social network data publishing,” in *Proc. IEEE Int. Conf. Big Knowl. (ICBK)*, Aug. 2017, pp. 208–213.

[132] S. H. Erfani and R. Mortazavi, “A novel graph-modiïňĄcation technique for user privacy-preserving on social networks,” *J. Telecommun. Inf. Technol.*, vol. 3, pp. 27–38, Oct. 2019.

[133] A. K. Rizi, M. N. Dehkordi, and N. N. Bakhsh, “SNI: Supervised anonymization technique to publish social networks having multiple sensitive labels,” *Secur. Commun. Netw.*, vol. 2019, pp. 1–23, Nov. 2019.

[134] M. Siddula, Y. Li, X. Cheng, Z. Tian, and Z. Cai, “Anonymization in online social networks based on enhanced equi-cardinal clustering,” *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 4, pp. 809–820, Aug. 2019.

[135] W. Shi, J. Hu, J. Yan, Z. Wu, and L. Lu, "A privacy measurement method using network structure entropy," in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, Oct. 2017, pp. 147–151.

[136] S. Mauw, Y. Ramírez-Cruz, and R. Trujillo-Rasua, "Conditional adjacency anonymity in social graphs under active attacks," *Knowl. Inf. Syst.*, vol. 61, no. 1, pp. 485–511, Oct. 2019.

[137] R. Mortazavi and S. H. Erfani, "An effective method for utility preserving social network graph anonymization based on mathematical modeling," *Int. J. Eng.*, vol. 31, no. 10, pp. 1624–1632, 2018.

[138] T. Wu, G. Ming, X. Xian, W. Wang, S. Qiao, and G. Xu, "Structural predictability optimization against inference attacks in data publishing," *IEEE Access*, vol. 7, pp. 92119–92136, 2019.

[139] J. Yan, L. Zhang, Y. Tian, G. Wen, and J. Hu, "An uncertain graph approach for preserving privacy in social networks based on important nodes," in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, Oct. 2018, pp. 107–111.

[140] D. Mohapatra and M. R. Patra, "Anonymization of attributed social graph using anatomy based clustering," *Multimedia Tools Appl.*, vol. 78, no. 18, pp. 25455–25486, Sep. 2019.

[141] K. Stokes, "Cover-up: A probabilistic privacy-preserving graph database model," *J. Ambient Intell. Humanized Comput.*, pp. 1–8, Oct. 2019, doi: 10.1007/s12652-019-01515-8.

[142] G. Wen, H. Liu, J. Yan, and Z. Wu, "A privacy analysis method to anonymous graph based on bayes rule in social networks," in *Proc. 14th Int. Conf. Comput. Intell. Secur. (CIS)*, Nov. 2018, pp. 469–472.

[143] F. Yu, M. Chen, B. Yu, W. Li, L. Ma, and H. Gao, "Privacy preservation based on clustering perturbation algorithm for social network," *Multimedia Tools Appl.*, vol. 77, no. 9, pp. 11241–11258, May 2018.

[144] M. Yuan, L. Chen, P. S. Yu, and T. Yu, "Protecting sensitive labels in social network data anonymization," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 3, pp. 633–647, Mar. 2013.

[145] J. Yuan, Y. Ou, and G. Gu, "An improved privacy protection method based on k-degree anonymity in social network," in *Proc. IEEE Int. Conf. Artif. Intell. Comput. Appl. (ICAICA)*, Mar. 2019, pp. 416–420.

[146] S. Hamzehzadeh and S. M. Mazinani, "ANNM: A new method for adding noise nodes which are used recently in anonymization methods in social networks," *Wireless Pers. Commun.*, vol. 107, no. 4, pp. 1995–2017, Aug. 2019.

[147] P. Liu, Y. Bai, L. Wang, and X. Li, "Partial k-Anonymity for privacy-preserving social network data publishing," *Int. J. Softw. Eng. Knowl. Eng.*, vol. 27, no. 1, pp. 71–90, Feb. 2017.

[148] X. Li, J. Yang, Z. Sun, and J. Zhang, "Publishing social graphs with differential privacy guarantees based on wPINQ," *Chin. J. Electron.*, vol. 28, no. 2, pp. 273–279, Mar. 2019.

[149] A.-T. Hoang, B. Carminati, and E. Ferrari, "Cluster-based anonymization of directed graphs," in *Proc. IEEE 5th Int. Conf. Collaboration Internet Comput. (CIC)*, Dec. 2019, pp. 91–100.

[150] H. Zhu, X. Zuo, and M. Xie, "DP-FT: A differential privacy graph generation with field theory for social network data release," *IEEE Access*, vol. 7, pp. 164304–164319, 2019.

[151] C. Kong, H. Li, H. Zhu, Y. Xiu, J. Liu, and T. Liu, "Anonymized user linkage under differential privacy," in *Proc. Int. Conf. Soft Comput. Data Sci.*, Iizuka, Japan: Springer, 2019, pp. 309–324.

[152] D. Wang and S. Long, "Boosting the accuracy of differentially private in weighted social networks," *Multimedia Tools Appl.*, vol. 78, no. 24, pp. 34801–34817, Dec. 2019.

[153] K. R. Macwan and S. J. Patel, "Node differential privacy in social graph degree publishing," *Procedia Comput. Sci.*, vol. 143, pp. 786–793, Jan. 2018.

[154] T. Gao and F. Li, "Sharing social networks using a novel differentially private graph model," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2019, pp. 1–4.

[155] B. P. Nguyen, H. Ngo, J. Kim, and J. Kim, "Publishing graph data with subgraph differential privacy," in *Proc. Int. Workshop Inf. Secur. Appl.* Jeju-do, South Korea: Springer, 2015, pp. 134–145.

[156] Q. Liu, G. Wang, F. Li, S. Yang, and J. Wu, "Preserving privacy with probabilistic indistinguishability in weighted social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 5, pp. 1417–1429, May 2017.

[157] A. Srivastava and G. Geethakumari, "Privacy preserving solution to prevent classification inference attacks in online social networks," *Int. J. Data Sci.*, vol. 4, no. 1, pp. 31–44, 2019.

[158] X. Ding, C. Wang, K.-K.-R. Choo, and H. Jin, "A novel privacy preserving framework for large scale graph data publishing," *IEEE Trans. Knowl. Data Eng.*, early access, Jul. 30, 2019, doi: 10.1109/TKDE.2019.2931903.

[159] Y. Guo, Z. Liu, Y. Zeng, R. Wang, and J. Ma, "Preserving privacy for hubs and links in social networks," in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, Oct. 2018, pp. 263–269.

[160] S. Mauw, Y. Ramírez-Cruz, and R. Trujillo-Rasua, "Anonymising social graphs in the presence of active attackers," *Trans. Data Priv.*, vol. 11, no. 2, pp. 169–198, 2018.

[161] X. Yin, S. Zhang, and H. Xu, "Node attributed query access algorithm based on improved personalized differential privacy protection in social network," *Int. J. Wireless Inf. Netw.*, vol. 26, no. 3, pp. 165–173, Sep. 2019.

[162] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Trans. Depend. Sec. Comput.*, vol. 15, no. 4, pp. 577–590, Aug. 2018.

[163] X. Zheng, G. Luo, and Z. Cai, "A fair mechanism for private data publication in online social networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 880–891, Apr. 2020.

[164] Y. Xie and M. Zheng, "A differentiated anonymity algorithm for social network privacy preservation," *Algorithms*, vol. 9, no. 4, p. 85, Dec. 2016.

[165] G. Cormode, D. Srivastava, T. Yu, and Q. Zhang, "Anonymizing bipartite graph data using safe groupings," *Proc. VLDB Endowment*, vol. 1, no. 1, pp. 833–844, 2008.

[166] H. Huang, D. Zhang, F. Xiao, K. Wang, J. Gu, and R. Wang, "Privacy-preserving approach PBCN in social network with differential privacy," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 2, pp. 931–945, Jun. 2020.

[167] T. Ji, C. Luo, Y. Guo, Q. Wang, L. Yu, and P. Li, "Community detection in online social networks: A differentially private and parsimonious approach," *IEEE Trans. Comput. Social Syst.*, vol. 7, no. 1, pp. 151–163, Feb. 2020.

[168] Y. Wang, J. Yang, and J. Zhang, "Differential privacy for weighted network based on probability model," *IEEE Access*, vol. 8, pp. 80792–80800, 2020.

[169] D. Al-Azizy, D. Millard, I. Symeonidis, K. O'Hara, and N. Shadbolt, "A literature survey and classifications on data deanonymisation," in *Proc. Int. Conf. Risks Secur. Internet Syst.* Hammamet, Tunisia: Springer, 2015, pp. 36–51.

[170] G. Beigi and H. Liu, "A survey on privacy in social media: Identification, mitigation, and applications," *ACM/IMS Trans. Data Sci.*, vol. 1, no. 1, pp. 1–38, Mar. 2020.

[171] D. Yin, Y. Shen, and C. Liu, "Attribute couplet attacks and privacy preservation in social networks," *IEEE Access*, vol. 5, pp. 25295–25305, 2017.

[172] Y. Li, Z. Su, J. Yang, and C. Gao, "Exploiting similarities of user friendship networks across social networks for user identification," *Inf. Sci.*, vol. 506, pp. 78–98, Jan. 2020.

[173] J. Mao, W. Tian, Y. Yang, and J. Liu, "An efficient social attribute inference scheme based on social links and attribute relevance," *IEEE Access*, vol. 7, pp. 153074–153085, 2019.

[174] C. Zhang, H. Jiang, Y. Wang, Q. Hu, J. Yu, and X. Cheng, "User identity de-anonymization based on attributes," in *Proc. Int. Conf. Wireless Algorithms, Syst., Appl.* Qingdao, China: Springer, 2019, pp. 458–469.

[175] C. Zhang, S. Wu, H. Jiang, Y. Wang, J. Yu, and X. Cheng, "Attribute-enhanced de-anonymization of online social networks," in *Int. Conf. Comput. Data Social Netw.* Dallas, TX, USA: Springer, 2019, pp. 256–267.

[176] Y. Li, Z. Zhang, Y. Peng, H. Yin, and Q. Xu, "Matching user accounts based on user generated content across social networks," *Future Gener. Comput. Syst.*, vol. 83, pp. 104–115, Jun. 2018.

[177] Y. Nie, Y. Jia, S. Li, X. Zhu, A. Li, and B. Zhou, "Identifying users across social networks based on dynamic core interests," *Neurocomputing*, vol. 210, pp. 107–115, Oct. 2016.

[178] Y. Sha, Q. Liang, and K. Zheng, "Matching user accounts across social networks based on users message," *Procedia Comput. Sci.*, vol. 80, pp. 2423–2427, Jan. 2016.

[179] L. Wang, K. Hu, Y. Zhang, and S. Cao, "Factor graph model based user profile matching across social networks," *IEEE Access*, vol. 7, pp. 152429–152442, 2019.

[180] W. Ahmad and R. Ali, ''Social account matching in online social media using cross-linked posts,'' *Procedia Comput. Sci.*, vol. 152, pp. 222–229, Jan. 2019.

[181] G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, ''A practical attack to de-anonymize social network users,'' in *Proc. IEEE Symp. Secur. Privacy*, May 2010, pp. 223–238.

[182] S. Nilizadeh, A. Kapadia, and Y.-Y. Ahn, ''Community-enhanced de-anonymization of online social networks,'' in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. - CCS*, 2014, pp. 537–548.

[183] W.-H. Lee, C. Liu, S. Ji, P. Mittal, and R. B. Lee, ''Blind de-anonymization attacks using social networks,'' in *Proc. Workshop Privacy Electron. Soc. - WPES*, 2017, pp. 1–4.

[184] J. Fang, A. Li, Q. Jiang, S. Li, and W. Han, ''A structure-based de-anonymization attack on graph data using weighted neighbor match,'' in *Proc. IEEE 4th Int. Conf. Data Sci. Cyberspace (DSC)*, Jun. 2019, pp. 480–486.

[185] S. Labitzke, F. Werling, J. Mittag, and H. Hartenstein, ''Do online social network friends still threaten my privacy?'' in *Proc. 3rd ACM Conf. Data Appl. Secur. Privacy - CODASPY*, 2013, pp. 13–24.

[186] Y. Qu, S. Yu, W. Zhou, and J. Niu, ''FBI: Friendship learning-based user identification in multiple social networks,'' in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[187] X. Gao, W. Ji, Y. Li, Y. Deng, and W. Dong, ''User identification with spatio-temporal awareness across social networks,'' in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 1831–1834.

[188] H. Yoshiura, ''Re-identifying people from anonymous histories of their activities,'' in *Proc. IEEE 10th Int. Conf. Awareness Sci. Technol. (iCAST)*, Oct. 2019, pp. 1–5.

[189] Y. Li, Y. Peng, Z. Zhang, M. Wu, Q. Xu, and H. Yin, ''A deep dive into user display names across social networks,'' *Inf. Sci.*, vol. 447, pp. 186–204, Jun. 2018.

[190] Y. Wang, C. Feng, L. Chen, H. Yin, C. Guo, and Y. Chu, ''User identity linkage across social networks via linked heterogeneous network embedding,'' *World Wide Web*, vol. 22, no. 6, pp. 2611–2632, Nov. 2019.

[191] D. Zhao, N. Zheng, M. Xu, X. Yang, and J. Xu, ''An improved user identification method across social networks via tagging behaviors,'' in *Proc. IEEE 30th Int. Conf. Tools with Artif. Intell. (ICTAI)*, Nov. 2018, pp. 616–622.

[192] U. Weinsberg, S. Bhagat, S. Ioannidis, and N. Taft, ''BlurMe: Inferring and obfuscating user gender based on ratings,'' in *Proc. 6th ACM Conf. Recommender Syst. - RecSys*, 2012, pp. 195–202.

[193] M. Kosinski, D. Stillwell, and T. Graepel, ''Private traits and attributes are predictable from digital records of human behavior,'' *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 15, pp. 5802–5805, Apr. 2013.

[194] A. Chaabane, G. Acs, and M. A. Kaafar, ''You are what you like! Information leakage through users' interests,'' in *Proc. 19th Annu. Netw. Distrib. Syst. Secur. Symp. (NDSS)*. San Diego, CA, USA: Citeseer, Feb. 2012, pp. 1–14.

[195] N. Z. Gong and B. Liu, ''Attribute inference attacks in online social networks,'' *ACM Trans. Privacy Secur.*, vol. 21, no. 1, pp. 1–30, Jan. 2018.

[196] H. Zhao, J. Chi, Y. Tian, and G. J. Gordon, ''Trade-offs and guarantees of adversarial representation learning for information obfuscation,'' 2019, *arXiv:1906.07902*. [Online]. Available: http://arxiv.org/abs/1906.07902

[197] P. Wang, J. Guo, Y. Lan, J. Xu, and X. Cheng, ''Your cart tells you: Inferring demographic attributes from purchase data,'' in *Proc. 9th ACM Int. Conf. Web Search Data Mining - WSDM*, 2016, pp. 173–182.

[198] H. Jiang, J. Yu, C. Hu, C. Zhang, and X. Cheng, ''SA framework based de-anonymization of social networks,'' *Procedia Comput. Sci.*, vol. 129, pp. 358–363, Jan. 2018.

[199] J. G. Thangam and A. Sankar, ''Emphasizing on space complexity in enterprise social networks for the investigation of link prediction using hybrid approach,'' in *Business Intelligence for Enterprise Internet of Things*. Cham, Switzerland: Springer, 2020, pp. 253–270.

[200] G. Song, Y. Zhou, H. Liu, G. Wen, and P. Ren, ''A privacy inference model based on attribute graph,'' in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, Oct. 2018, pp. 97–101.

[201] X. Xian, T. Wu, S. Qiao, W. Wang, Y. Liu, and N. Han, ''Multi-view low-rank coding-based network data de-anonymization,'' *IEEE Access*, vol. 8, pp. 94575–94593, 2020.

[202] P. Gundecha, G. Barbier, and H. Liu, ''Exploiting vulnerability to secure user privacy on a social networking site,'' in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining - KDD*, 2011, pp. 511–519.

[203] B. A. Pijani, A. Imine, and M. Rusinowitch, ''Inferring attributes with picture metadata embeddings,'' *ACM SIGAPP Appl. Comput. Rev.*, vol. 20, no. 2, pp. 36–45, Jul. 2020.

[204] H. Li, Q. Chen, H. Zhu, D. Ma, H. Wen, and X. Shen, ''Privacy leakage via de-anonymization and aggregation in heterogeneous social networks,'' *IEEE Trans. Depend. Sec. Comput.*, vol. 17, no. 2, pp. 350–362, Mar. 2020.

[205] L. Guo, C. Zhang, and Y. Fang, ''A trust-based privacy-preserving friend recommendation scheme for online social networks,'' *IEEE Trans. Depend. Sec. Comput.*, vol. 12, no. 4, pp. 413–427, Jul. 2015.

[206] V. B. Kukkala and S. R. S. Iyengar, ''Identifying influential spreaders in a social network (While preserving Privacy),'' *Proc. Privacy Enhancing Technol.*, vol. 2020, no. 2, pp. 537–557, Apr. 2020.

[207] Q. Dong and D. Huang, ''Privacy-preserving matchmaking in geosocial networks with untrusted servers,'' in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 2591–2592.

[208] S. Zhang, Q. Liu, and Y. Lin, ''Anonymizing popularity in online social networks with full utility,'' *Future Gener. Comput. Syst.*, vol. 72, pp. 227–238, Jul. 2017.

[209] T. Georgiou, A. El Abbadi, and X. Yan, ''Privacy-preserving community-aware trending topic detection in online social media,'' in *Proc. IFIP Annu. Conf. Data Appl. Secur. Privacy*. Regensburg, Germany: Springer, 2017, pp. 205–224.

[210] C. Valliyammai and A. Bhuvaneswari, ''Semantics-based sensitive topic diffusion detection framework towards privacy aware online social networks,'' *Cluster Comput.*, vol. 22, no. S1, pp. 407–422, Jan. 2019.

[211] J.-R. Gao, W. Chen, J.-J. Xu, A. Liu, Z.-X. Li, H. Yin, and L. Zhao, ''An efficient framework for multiple subgraph pattern matching models,'' *J. Comput. Sci. Technol.*, vol. 34, no. 6, pp. 1185–1202, Nov. 2019.

[212] J. Casas-Roma, ''DUEF-GA: Data utility and privacy evaluation framework for graph anonymization,'' *Int. J. Inf. Secur.*, vol. 19, no. 4, pp. 465–478, Aug. 2020.

[213] D. Li, Q. Lv, L. Shang, and N. Gu, ''Efficient privacy-preserving content recommendation for online social communities,'' *Neurocomputing*, vol. 219, pp. 440–454, Jan. 2017.

[214] A. Mosallanezhad, G. Beigi, and H. Liu, ''Deep reinforcement learning-based text anonymization against private-attribute inference,'' in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 2360–2369.

[215] T. Gao and F. Li, ''Privacy-preserving sketching for online social network data publication,'' in *Proc. 16th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2019, pp. 1–9.

[216] Y. Liu, J. Liu, Z. Zhang, L. Zhu, and A. Li, ''Rem: From structural entropy to community structure deception,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 12938–12948.

[217] S. Ramezanian and V. Niemi, ''Privacy preserving cyberbullying prevention with AI methods in 5G networks,'' in *Proc. 25th Conf. Open Innov. Assoc. (FRUCT)*, Nov. 2019, pp. 265–271.

[218] S. Kavianpour, A. Tamimi, and B. Shanmugam, ''A privacy-preserving model to control social interaction behaviors in social network sites,'' *J. Inf. Secur. Appl.*, vol. 49, Dec. 2019, Art. no. 102402.

[219] N. C. Rathore and S. Tripathy, ''AppMonitor: Restricting information leakage to third-party applications,'' *Social Netw. Anal. Mining*, vol. 10, no. 1, pp. 1–20, Dec. 2020.

[220] S. T. Boshrooyeh, A. Küpçü, and Ö. Özkasap, ''Privado: Privacy-preserving group-based advertising using multiple independent social network providers,'' *ACM Trans. Privacy Secur.*, vol. 23, no. 3, pp. 1–36, Jul. 2020.

[221] Y. Huang, Z. Cai, and A. G. Bourgeois, ''Privacy protection for context-aware services: A two-layer three-party game model,'' in *Proc. Int. Conf. Wireless Algorithms, Syst., Appl.* Qingdao, China: Springer, 2019, pp. 124–136.

[222] L. Guo, C. Zhang, Y. Fang, and P. Lin, ''A privacy-preserving attribute-based reputation system in online social networks,'' *J. Comput. Sci. Technol.*, vol. 30, no. 3, pp. 578–597, May 2015.

[223] J. Yang, X. Li, Z. Sun, and J. Zhang, ''A differential privacy framework for collaborative filtering,'' *Math. Problems Eng.*, vol. 2019, pp. 1–11, Jan. 2019.

[224] R. Aljably, Y. Tian, M. Al-Rodhaan, and A. Al-Dhelaan, ''Anomaly detection over differential preserved privacy in online social networks,'' *PLoS ONE*, vol. 14, no. 4, Apr. 2019, Art. no. e0215856.

[225] P. Li, F. Zhou, Z. Xu, Y. Li, and J. Xu, "Privacy-preserving graph operations for social network analysis," in *Proc. Int. Symp. Secur. Privacy Social Netw. Big Data*. Tianjin, China: Springer, 2020, pp. 303–317.

[226] X. Li, Y. Xin, C. Zhao, Y. Yang, and Y. Chen, "Graph convolutional networks for privacy metrics in online social networks," *Appl. Sci.*, vol. 10, no. 4, p. 1327, Feb. 2020.

[227] J. Alemany, E. Del Val, J. M. Alberola, and A. Garcia-Fornes, "Metrics for privacy assessment when sharing information in online social networks," *IEEE Access*, vol. 7, pp. 143631–143645, 2019.

[228] R. G. Pensa, "Enhancing privacy awareness in online social networks: A knowledge-driven approach," in *Proc. Int. Workshop Knowl.-Driven Anal. Impacting Hum. Quality Life (KDAH)*, vol. 2482, 2019, pp. 1–2.

[229] X. Li, Y. Xin, C. Zhao, Y. Yang, S. Luo, and Y. Chen, "Using user behavior to measure privacy on online social networks," *IEEE Access*, vol. 8, pp. 108387–108401, 2020.

[230] R. G. Pensa, G. Di Blasi, and L. Bioglio, "Network-aware privacy risk estimation in online social networks," *Social Netw. Anal. Mining*, vol. 9, no. 1, p. 15, Dec. 2019.

[231] F. Zhou, K. Zhang, S. Xie, and X. Luo, "Learning to correlate accounts across online social networks: An embedding-based approach," *Informs J. Comput.*, vol. 32, no. 3, pp. 714–729, Jul. 2020.

[232] I. Nurgaliev, Q. Qu, S. M. H. Bamakan, and M. Muzammal, "Matching user identities across social networks with limited profile data," *Frontiers Comput. Sci.*, vol. 14, no. 6, pp. 1–14, Dec. 2020.

[233] X. Han, H. Huang, and L. Wang, "F-PAD: Private attribute disclosure risk estimation in online social networks," *IEEE Trans. Depend. Sec. Comput.*, vol. 16, no. 6, pp. 1054–1069, Nov. 2019.

[234] G. Xu, H. Li, Y. Zhang, S. Xu, J. Ning, and R. Deng, "Privacy-preserving federated deep learning with irregular users," *IEEE Trans. Depend. Sec. Comput.*, early access, Jun. 30, 2020, doi: 10.1109/TDSC.2020.3005909.

[235] R. Hu, Y. Guo, H. Li, Q. Pei, and Y. Gong, "Personalized federated learning with differential privacy," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9530–9539, Oct. 2020.

[236] X. Liu, H. Li, G. Xu, R. Lu, and M. He, "Adaptive privacy-preserving federated learning," *Peer Peer Netw. Appl.*, vol. 13, no. 6, pp. 2356–2366, Nov. 2020.

[237] A. Triastcyn and B. Faltings, "Federated generative privacy," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 50–57, Jul. 2020.

[238] O. Choudhury, A. Gkoulalas-Divanis, T. Salonidis, I. Sylla, Y. Park, G. Hsu, and A. Das, "Anonymizing data for privacy-preserving federated learning," 2020, *arXiv:2002.09096*. [Online]. Available: http://arxiv.org/abs/2002.09096

[239] X. Hu, L.-E. Wang, J. Tang, C. Lei, P. Liu, and X. Li, "Anonymizing approach to resist label-neighborhood attacks in dynamic releases of social networks," in *Proc. IEEE 19th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Oct. 2017, pp. 1–6.

[240] R. Yue, Y. Li, T. Wang, and Y. Jin, "An efficient adaptive graph anonymization framework for incremental data publication," in *Proc. 5th Int. Conf. Behav., Econ., Socio-Cultural Comput. (BESC)*, Nov. 2018, pp. 103–108.

[241] C.-H. Tai, P.-J. Tseng, P. S. Yu, and M.-S. Chen, "Identity protection in sequential releases of dynamic networks," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 3, pp. 635–651, Mar. 2014.

[242] W. M. Yafooz, Z. B. A. Bakar, S. A. Fahad, and A. M. Mithon, "Business intelligence through big data analytics, data mining and machine learning," in *Data Management, Analytics and Innovation*. Singapore: Springer, 2020, pp. 217–230.

[243] J. Liu, X. Li, L. Ye, H. Zhang, X. Du, and M. Guizani, "BPDS: A blockchain based privacy-preserving data sharing for electronic medical records," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[244] Y. Khazbak, J. Qiu, T. Tan, and G. Cao, "TargetFinder: A privacy preserving system for locating targets through IoT cameras," *ACM Trans. Internet Things*, vol. 1, no. 3, pp. 1–23, Jul. 2020.

[245] M. Shen, G. Cheng, L. Zhu, X. Du, and J. Hu, "Content-based multi-source encrypted image retrieval in clouds with privacy preservation," *Future Gener. Comput. Syst.*, vol. 109, pp. 621–632, Aug. 2020.

[246] R. Wei, H. Shen, and H. Tian, "An improved (k,p,l)-anonymity method for privacy preserving collaborative filtering," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.

[247] A. Bhatia and S. Shaikh, "Proceedings for the first international workshop on social threats in online conversations: Understanding and management," in *Proc. 1st Int. Workshop Social Threats Online Conversations, Understand. Manage.*, May 2020, pp. 1–6.

[248] R. Talat, M. S. Obaidat, M. Muzammal, A. H. Sodhro, Z. Luo, and S. Pirbhulal, "A decentralised approach to privacy preserving trajectory mining," *Future Gener. Comput. Syst.*, vol. 102, pp. 382–392, Jan. 2020.

[249] H. Wang, X. A. Wang, S. Xiao, and J. Liu, "Decentralized data outsourcing auditing protocol based on blockchain," *J. Ambient Intell. Humanized Comput.*, pp. 1–12, Aug. 2020, doi: 10.1007/s12652-020-02432-x.

[250] M. A. P. Chamikara, P. Bertok, D. Liu, S. Camtepe, and I. Khalil, "Efficient privacy preservation of big data for accurate data mining," *Inf. Sci.*, vol. 527, pp. 420–443, Jul. 2020.

[251] P. Zhao, H. Huang, X. Zhao, and D. Huang, "p^3: Privacy-preserving scheme against poisoning attacks in mobile-edge computing," *IEEE Trans. Comput. Social Syst.*, vol. 7, no. 3, pp. 818–826, Jun. 2020.

[252] Y. Zhao, J. Zhao, L. Jiang, R. Tan, D. Niyato, Z. Li, L. Lyu, and Y. Liu, "Privacy-preserving blockchain-based federated learning for IoT devices," *IEEE Internet Things J.*, early access, Aug. 18, 2020, doi: 10.1109/JIOT.2020.3017377.

[253] S. Madan and P. Goswami, "A privacy preservation model for big data in map-reduced framework based on k-anonymisation and swarm-based algorithms," *Int. J. Intell. Eng. Inform.*, vol. 8, no. 1, pp. 38–53, 2020.

[254] R. Jain, N. Jain, and A. Nayyar, "Security and privacy in social networks: Data and structural anonymity," in *Handbook of Computer Networks and Cyber Security*. Singapore: Springer, 2020, pp. 265–293.

[255] C. Fang, Y. Guo, N. Wang, and A. Ju, "Highly efficient federated learning with strong privacy preservation in cloud computing," *Comput. Secur.*, vol. 96, Sep. 2020, Art. no. 101889.

[256] S. D. Khambalkar, S. D. Kamble, N. V. Thakur, N. U. Sambhe, and N. S. Mangrulkar, "An overview on privacy preservation and public auditing on outsourced cloud data," in *Smart Trends in Computing and Communications*. Singapore: Springer, 2020, pp. 463–470.

[257] J. Li, X. Kuang, S. Lin, X. Ma, and Y. Tang, "Privacy preservation for machine learning training and classification based on homomorphic encryption schemes," *Inf. Sci.*, vol. 526, pp. 166–179, Jul. 2020.

[258] Y.-L. Gao, X.-B. Chen, G. Xu, W. Liu, M.-X. Dong, and X. Liu, "A new blockchain-based personal privacy protection scheme," *Multimedia Tools Appl.*, pp. 1–14, Sep. 2020, doi: 10.1007/s11042-020-09867-6.

[259] C. Butpheng, K.-H. Yeh, and H. Xiong, "Security and privacy in IoT-cloud-based e-health systems—A comprehensive review," *Symmetry*, vol. 12, no. 7, p. 1191, 2020.

[260] N. Kaaniche and M. Laurent, "Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms," *Comput. Commun.*, vol. 111, pp. 120–141, Oct. 2017.

[261] E. Damiani, C. A. Ardagna, F. Zavatarelli, E. Rekleitis, and L. Marinos, "Big data threat landscape and good practice guide," in *European Union Agency For Network and Information Security*. Heraklion, Crete: EU Agency for Network and Information Security (ENISA), 2016.

[262] A. Ghorbel, M. Ghorbel, and M. Jmaiel, "Privacy in cloud computing environments: A survey and research challenges," *J. Supercomput.*, vol. 73, no. 6, pp. 2763–2800, Jun. 2017.

[263] H. F. Atlam and G. B. Wills, "Iot security, privacy, safety and ethics," in *Digital Twin Technologies and Smart Cities*. Cham, Switzerland: Springer, 2020, pp. 123–149.

[264] J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, and R. Buyya, "Ensuring security and privacy preservation for cloud data services," *ACM Comput. Surv.*, vol. 49, no. 1, pp. 1–39, Jul. 2016.

[265] M. Sumathi and S. Sangeetha, "A group-key-based sensitive attribute protection in cloud storage using modified random fibonacci cryptography," *Complex Intell. Syst.*, pp. 1–15, Jun. 2020, doi: 10.1007/s40747-020-00162-3.

[266] P. J. Sun, "The optimal privacy strategy of cloud service based on evolutionary game," *Cluster Comput.*, pp. 1–19, Aug. 2020, doi: 10.1007/s10586-020-03164-5.

[267] P. Sun, "Security and privacy protection in cloud computing: Discussions and challenges," *J. Netw. Comput. Appl.*, vol. 160, Jun. 2020, Art. no. 102642.

[268] U. Khadam, M. M. Iqbal, M. Alruily, M. A. Al Ghamdi, M. Ramzan, and S. H. Almotiri, "Text data security and privacy in the Internet of Things: Threats, challenges, and future directions," *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 1–15, Feb. 2020.

[269] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, "PPaaS: Privacy preservation as a service," 2020, *arXiv:2007.02013*. [Online]. Available: http://arxiv.org/abs/2007.02013

[270] A. Ye, Q. Zhang, Y. Diao, J. Zhang, H. Deng, and B. Cheng, "A semantic-based approach for privacy-preserving in trajectory publishing," *IEEE Access*, vol. 8, pp. 184965–184975, 2020.

[271] C.-Y. Lin, "Suppression techniques for privacy-preserving trajectory data publishing," *Knowl.-Based Syst.*, vol. 206, Oct. 2020, Art. no. 106354.

[272] N. Kaaniche, M. Laurent, and S. Belguith, "Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey," *J. Netw. Comput. Appl.*, vol. 171, Dec. 2020, Art. no. 102807.

[273] D. H. Ramírez and J. M. Auñón, "Privacy preserving K-Means clustering: A secure multi-party computation approach," 2020, *arXiv:2009.10453*. [Online]. Available: http://arxiv.org/abs/2009.10453

[274] N. Lisin and S. Zapechnikov, "Methods and approaches for privacy-preserving machine learning," in *Advanced Technologies in Robotics and Intelligent Systems*. Cham, Switzerland: Springer, 2020, pp. 141–148.

[275] Y. Huang, M. Milani, and F. Chiang, "PACAS: Privacy-aware, data Cleaning-as-a-Service," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 1023–1030.

[276] Y. Huang, M. Milani, and F. Chiang, "Privacy-aware data cleaning-as-a-service," *Inf. Syst.*, vol. 94, Dec. 2020, Art. no. 101608.

[277] H. Y. Youm, "An overview of de-identification techniques and their standardization directions," *IEICE Trans. Inf. Syst.*, vol. E103.D, no. 7, pp. 1448–1461, 2020.

[278] J. Hiscott, M. Alexandridi, M. Muscolini, E. Tassone, E. Palermo, M. Soultsioti, and A. Zevini, "The global impact of the coronavirus pandemic," *Cytokine Growth Factor Rev.*, vol. 53, pp. 1–9, Jun. 2020, doi: 10.1016/j.cytogfr.2020.05.010.

[279] I. Y. Iourov, M. A. Zelenova, and S. G. Vorsanova, "Covid-19: A crash test for biomedical publishing," *MedRxiv*, pp. 1–10, Aug. 2020, doi: 10.1101/2020.06.13.20130310.

[280] M. Shen, Y. Wei, and T. Li, "Bluetooth-based COVID-19 proximity tracing proposals: An overview," 2020, *arXiv:2008.12469*. [Online]. Available: http://arxiv.org/abs/2008.12469

[281] N. Y. Ahn, J. E. Park, D. H. Lee, and P. C. Hong, "Balancing personal privacy and public safety during COVID-19: The case of South Korea," 2020, *arXiv:2004.14495*. [Online]. Available: http://arxiv.org/abs/2004.14495

[282] J. Budd, B. S. Miller, E. M. Manning, V. Lampos, M. Zhuang, M. Edelstein, G. Rees, V. C. Emery, M. M. Stevens, N. Keegan, and M. J. Short, "Digital technologies in the public-health response to COVID-19," *Nature Med.*, vol. 26, no. 8, pp. 1183–1192, 2020.

[283] M. Nanni *et al.*, "Give more data, awareness and control to individual citizens, and they will help COVID-19 containment," 2020, *arXiv:2004.05222*. [Online]. Available: http://arxiv.org/abs/2004.05222

[284] A. Sardi, A. Rizzi, E. Sorano, and A. Guerrieri, "Cyber risk in health facilities: A systematic literature review," *Sustainability*, vol. 12, no. 17, p. 7002, Aug. 2020.

[285] A. Apostolos and K. Apostolos, "Tracking applications: A factor of mithridatism of personal data and privacy in the Post-COVID-19 era," *Disaster Med. Public Health Preparedness*, vol. 14, no. 3, p. e27, Jun. 2020.

[286] M. Moeini, M. Möhlmann, and J. Hummel, "Understanding knotted tensions in purveying pandemic public monitoring technologies," *SSRN Electron. J.*, pp. 1–36, Aug. 2020, doi: 10.2139/ssrn.3671458.

[287] A. Zwitter and O. J. Gstrein, "Big data, privacy and COVID-19—Learning from humanitarian expertise in data protection," *Int. J. Humanitarian Action*, vol. 5, no. 4, 2020, doi: 10.1186/s41018-020-00072-6.

[288] A. B. Dar, A. H. Lone, S. Zahoor, A. A. Khan, and R. Naaz, "Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions," *Comput. Sci. Rev.*, vol. 38, Nov. 2020, Art. no. 100307.

[289] A. V. Lucca, R. Luchtenberg, L. G. de Paula Conceicao, L. A. Silva, R. G. Ovejero, M. Navarro-Cáceres, and V. R. Q. Leithardt, "System for control and management of data privacy of patients with COVID-19," to be published, doi: 10.20944/preprints202007.0369.v1.

[290] Z. Allam, "Actualizing big data through revised data protocols to render more accurate infectious disease monitoring and modeling," *Surveying COVID-19 Pandemic Implications*, pp. 71–79, Jul. 2020, doi: 10.1016/B978-0-12-824313-8.00004-8.

[291] V. Shubina, S. Holcer, M. Gould, and E. S. Lohan, "Survey of decentralized solutions with mobile devices for user location tracking, proximity detection, and contact tracing in the COVID-19 era," *Data*, vol. 5, no. 4, p. 87, Sep. 2020.

[292] K. Matsui, K. Yamamoto, and Y. Inoue, "Professional commitment to ethical discussions needed from epidemiologists in the COVID-19 pandemic," *J. Epidemiology*, vol. 30, no. 9, pp. 375–376, 2020.

[293] J. Zhang and M. Wu, "Blockchain use in IoT for privacy-preserving anti-pandemic home quarantine," *Electronics*, vol. 9, no. 10, p. 1746, Oct. 2020.

[294] A. Verri Lucca, L. Augusto Silva, R. Luchtenberg, L. Garcez, X. Mao, R. García Ovejero, I. Miguel Pires, J. Luis Victória Barbosa, and V. Reis Quietinho Leithardt, "A case study on the development of a data privacy management solution based on patient information," *Sensors*, vol. 20, no. 21, p. 6030, Oct. 2020.

[295] D. Gao, Y. Liu, A. Huang, C. Ju, H. Yu, and Q. Yang, "Privacy-preserving heterogeneous federated transfer learning," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 2552–2559.

[296] H. C. Tanuwidjaja, R. Choi, and K. Kim, "A survey on deep learning techniques for privacy-preserving," in *Proc. Int. Conf. Mach. Learn. Cyber Secur.*, Guangzhou, China: Springer, 2019, pp. 29–46.

[297] C. Chen, B. Wu, M. Qiu, L. Wang, and J. Zhou, "A comprehensive analysis of information leakage in deep transfer learning," 2020, *arXiv:2009.01989*. [Online]. Available: http://arxiv.org/abs/2009.01989

[298] M. Grama, M. Musat, L. Muñoz-González, J. Passerat-Palmbach, D. Rueckert, and A. Alansary, "Robust aggregation for adaptive privacy preserving federated learning in healthcare," 2020, *arXiv:2009.08294*. [Online]. Available: http://arxiv.org/abs/2009.08294

[299] K. M. Hossein, M. E. Esmaeili, T. Dargahi, and A. Khonsari, "Blockchain-based privacy-preserving healthcare architecture," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, May 2019, pp. 1–4.

[300] Y. Xie, H. Wang, B. Yu, and C. Zhang, "Secure collaborative few-shot learning," *Knowl.-Based Syst.*, vol. 203, Sep. 2020, Art. no. 106157.

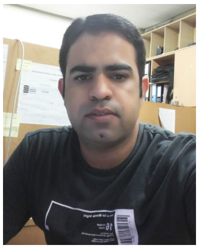
[301] Y. Zhou, S. Zheng, and L. Wang, "Privacy-preserving and efficient public key encryption with keyword search based on CP-ABE in cloud," *Cryptography*, vol. 4, no. 4, p. 28, Oct. 2020.

[302] X. Ma, C. Wang, and L. Wang, "The data sharing scheme based on blockchain," in *Proc. 2nd ACM Int. Symp. Blockchain Secure Crit. Infrastruct.*, Oct. 2020, pp. 96–105.

[303] W. Xiong and L. Xiong, "Data resource protection based on smart contract," *Comput. Secur.*, vol. 98, Nov. 2020, Art. no. 102004.

[304] S. Zeadally, E. Adi, Z. Baig, and I. A. Khan, "Harnessing artificial intelligence capabilities to improve cybersecurity," *IEEE Access*, vol. 8, pp. 23817–23837, 2020.

[305] R. Wang and W. Ji, "Computational intelligence for information security: A survey," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 4, no. 5, pp. 616–629, Oct. 2020.



ABDUL MAJEED received the B.Sc. degree in information technology from the University Institute of Information Technology (UIIT), PMAS-UAAR, Rawalpindi, Pakistan, in 2013, and the M.S. degree in information security from COMSATS University, Islamabad, Pakistan, in 2016. He is currently pursuing the Ph.D. degree with Korea Aerospace University, Goyang, South Korea. He worked as a Security Analyst with Trillium Information Security Systems (TISS), Rawalpindi, from 2015 to 2016. His research interests include privacy preserving data publishing (PPDP), information security, robotics, data mining, social network analysis and mining, and machine learning.

SUNGCHANG LEE (Member, IEEE) received the B.S. degree from Kyungpook National University, in 1983, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), in 1985, and the Ph.D. degree in electrical engineering from Texas A&M University, in 1991. From 1985 to 1987, he was with KAIST, as a Researcher, where he worked on Image Processing and Pattern Recognition Projects. From 1992 to 1993, he was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI), South Korea. From 2004 to 2009, he was the Director of the Government Project on Intelligent Smart Home Security and Automation Service Technology. In 2009, he was the Vice President of the Institute of Electronics and Information Engineers (IEIE), South Korea, and also the Director of the Telecommunications Society, South Korea. Since 1993, he has been a Faculty with Korea Aerospace University, Goyang, South Korea, where he is currently a Professor with the School of Electronics, Telecommunication and Computer Engineering.