A Survey on Privacy in Social Media: Identification, Mitigation, and Applications

GHAZALEH BEIGI and HUAN LIU, Arizona State Univesity, USA

The increasing popularity of social media has attracted a huge number of people to participate in numerous activities on a daily basis. This results in tremendous amounts of rich user-generated data. These data provide opportunities for researchers and service providers to study and better understand users' behaviors and further improve the quality of the personalized services. Publishing user-generated data risks exposing individuals' privacy. Users privacy in social media is an emerging research area and has attracted increasing attention recently. These works study privacy issues in social media from the two different points of views: identification of vulnerabilities and mitigation of privacy risks. Recent research has shown the vulnerability of user-generated data against the two general types of attacks, identity disclosure and attribute disclosure. These privacy issues mandate social media data publishers to protect users' privacy by sanitizing user-generated data before publishing it. Consequently, various protection techniques have been proposed to anonymize user-generated social media data. There is vast literature on privacy of users in social media from many perspectives. In this survey, we review the key achievements of user privacy in social media. In particular, we review and compare the state-of-the-art algorithms in terms of the privacy leakage attacks and anonymization algorithms. We overview the privacy risks from different aspects of social media and categorize the relevant works into five groups: (1) social graphs and privacy, (2) authors in social media and privacy, (3) profile attributes and privacy, (4) location and privacy, and (5) recommendation systems and privacy. We also discuss open problems and future research directions regarding user privacy issues in social media.

CCS Concepts: • **Security and privacy → Data anonymization and sanitization**; **Social network security and privacy**; *Privacy protections*; • **Human-centered computing → Social networks**;

Additional Key Words and Phrases: Privacy protection, identification of vulnerabilities, mitigation of risks, social media

ACM Reference format:
Ghazaleh Beigi and Huan Liu. 2020. A Survey on Privacy in Social Media: Identification, Mitigation, and Applications. *ACM/IMS Trans. Data Sci.* 1, 1, Article 7 (January 2020), 38 pages.
https://doi.org/10.1145/3343038

1 INTRODUCTION

The explosive Web growth in the past decade has drastically changed the way billions of people all around the globe conduct numerous activities such as surfing the web, creating online profiles in social media platforms, interacting with other people, and sharing posts and various personal

This material is based upon the work supported in part by Army Research Office (ARO) under grant number W911NF-15-1-0328 and Office of Naval Research (ONR) under grant number N00014-17-1-2605.
Authors' address: G. Beigi and H. Liu, School of Computing, Informatics, and Decision Systems Engineering, Ira A. Fulton Schools of Engineering, Arizona State Univesity, Tempe, AZ, USA; emails: {gbeigi, huan.liu}@asu.edu.

2577-3224/2020/01-ART7 $15.00
https://doi.org/10.1145/3343038

information in a rich environment. This results in tremendous amounts of user-generated data. The massive amounts of user information and the availability of up-to-date data makes social media platforms an attractive target for organizations seeking to collect and aggregate this information either for legitimate purposes or nefarious goals [35]. For example, the user-generated data provide opportunities for researchers and business partners to study and understand individuals at unprecedented scales [19, 28]. This information is also crucial for online vendors to provide personalized services, and a lack of it would result in a deteriorating quality of online personalization service [23].

On the other hand, tremendous amounts of user-generated data risk exposing individuals' privacy due to its richness of content including a user's relationships and other private information [22, 26, 85, 140]. These data also make online users traceable, and, accordingly, users become severely vulnerable to potential risks ranging from persecution by governments to targeted fraud. For example, users may share their vacation plans publicly on Twitter without knowing that this information could be used by adversaries for break-ins and thefts in the future [124, 191]. Moreover, sensitive information that users do not usually explicitly disclose can be easily inferred from their activities in social media such as location [109, 123], age [178], and trust/distrust relationships [27, 29, 30].

Privacy issues could be prominent when the data get published by a data publisher or service provider. In general, two types of information disclosures have been identified in the literature: identity disclosure and attribute disclosure attacks [51, 103, 107]. Identity disclosure occurs when an individual is mapped to an instance in a released dataset. Attribute disclosure happens when the adversary could infer some new information regarding an individual based on the released data. Attribute disclosure becomes more probable when there is accurate disclosure of people's identities. Similarly, privacy leakage attacks in social media could be also categorized into either identity disclosure or attribute disclosure. These user privacy issues mandate social media data publishers to protect users' privacy by sanitizing user-generated data before they are published publicly.

Data anonymization is a complex problem, and its goal is to remove or perturb data to prevent adversaries from inferring sensitive information while ensuring the utility of the published data. One straightforward anonymization technique is to remove "Personally Identifiable Information" (a.k.a. PII) such as names, user ID, age, and location information. This solution has been shown to be far from sufficient in preserving privacy [19, 139]. An example of this insufficient approach is the anonymized dataset published for the Netflix prize challenge. As a part of the Netflix prize contest, Netflix publicly released a dataset containing movie ratings of 500,000 subscribers. The data were supposed to be anonymized, and all PII is removed from it. Narayanan et al. [139] propose a de-anonymization attack that map users' records in the anonymized dataset to corresponding profiles on IMDB. In particular, the results of this work show that the structure of the data carries enough information for a potential breach of privacy to re-identify anonymized users.

Consequently, various protection techniques have been proposed to anonymize user-generated social media data. In general, the ultimate goal of an anonymization approach is to preserve social media user privacy while ensuring the utility of published data. As a counterpart to this research direction, another group of works investigate the potential privacy breaches from social media user data by introducing new attacks. These works find the gaps in anonymizing user-generated data and further improve anonymization techniques.

There is vast literature on privacy of users in social media from many perspectives. Existing works cover three applications in social media, i.e., making connection with people, sharing contextual information, and receiving personalized services. Besides, users generate various types of data, including graph data, textual data, spatiotemporal data, and profile attribute data. This results in 12 pairs of applications and data type combination. We categorize existing works into

Table 1. Combination of Different Data Types and Applications Are Covered with Five Categories

		Applications		
		Making Connection with People	Sharing Contextual Information	Receiving Personalized Services
Type of Data	Graph data	• Social graphs & privacy • Profile attributes & privacy	• Profile attributes & privacy	• Recommendation systems & privacy
	Textual Data	• Profile attributes & privacy	• Authors & privacy • Profile attributes & privacy	• Recommendation systems & privacy
	Spatiotemporal Data	• Users location & privacy	• Users location & privacy	• Recommendation systems & privacy
	Profile Attribute Data	• Profile attributes & privacy	• Profile attributes & privacy	• Recommendation systems & privacy

five distinct categories to cover different combinations, including (1) social graphs and privacy, (2) authors in social media and privacy, (3) profile attributes and privacy, (4) location and privacy, and (5) recommendation systems and privacy. Table 1 shows how each category covers different combination of applications and data types. The goal of this article is to provide a comprehensive review of existing works on user privacy issues and solutions in social media and give a guidance on future research directions. The contributions of this survey are summarized as follows:

- We give an overview of the traditional privacy models for structured data and discuss how these models are adopted for privacy issues in social media. We formally define two types of privacy leakage disclosures that covers most of the existing definitions in the literature.
- We categorize privacy issues and solutions on social media into different groups: (1) social graphs and privacy, (2) authors in social media and privacy, (3) profile attributes and privacy, (4) location and privacy, and (5) recommendation systems and privacy. We overview existing works in each group with a principled way to group representative methods into different categories.
- We discuss several open issues and provide future directions for privacy in social media.

The remainder of this survey is organized as follows. In Section 2, we present an overview of traditional methods and formally define two types of privacy disclosures. In Section 3, we review the state-of-the-art methods for privacy of social media graphs. More specifically, Section 3.1. covers de-anonymization attacks in social graphs, and Section 3.2. covers anonymization techniques that are proposed for preserving privacy of graph data against de-anonymization attacks. We review author identification works in Section 4. In Sections 5 and 6, we overview state-of-the-art de-anonymization techniques for inferring users profile attributes and location information. In Section 7, privacy issues and solutions in recommendation systems are reviewed. Finally, we conclude this article in Section 8 by discussing the open issues and future directions.

2 TRADITIONAL PRIVACY MODELS

Privacy-preserving techniques were first introduced for tabular and micro data. With the emergence of social media, the issue of online user privacy was raised. Researchers then focus on studying privacy leakage issues as well as anonymization and privacy-preserving techniques specialized for social media data. There are two types of information disclosure in the literature: identity

disclosure and attribute disclosure attacks [51, 103, 107]. We can formally define these attacks as follows:

Definition 2.1 (Identity Disclosure Attack). Given $T = (\mathbf{G}, \mathbf{A}, \mathbf{B})$, which is a snapshot of a social media platform with a social graph $\mathbf{G} = (V, E)$ where V is the set of users and E demonstrates the social relations between them, a user behavior $\mathbf{A}$ and an attribute information $\mathbf{B}$, the identity disclosure attack is to map all users in the list of target users V_t to their known identities. For each $v \in V_t$, we have the information of her social friends and behavior.

Definition 2.2 (Attribute Disclosure Attack). Given $T = (\mathbf{G}, \mathbf{A}, \mathbf{B})$, which is a snapshot of a social media platform with a social graph $\mathbf{G} = (V, E)$, where V is the set of users and E demonstrates the social relations between them, a user behavior $\mathbf{A}$ and an attribute information $\mathbf{B}$, the attribute disclosure attack is used to infer the attributes a_v for all $v \in V_t$ where V_t is a list of targeted users. For each $v \in V_t$, we have the information of her social friends and behavior.

Network graph de-anonymization and author identification are examples of identity disclosure attacks that exists in social media. Examples of attribute disclosure attack include the disclosure of users' profile attributes, location, and preferences information in recommendation systems.

Before we discuss privacy leakage in social media, we overview the traditional privacy models for structured data such as k-anonymity [171], l-diversity [119], t-closeness [107], and differential privacy [52]. These models are defined over structured databases and cannot be directly applied to unstructured user generated data. The reason is that quasi-identifiers and sensitive attributes are not clear in the context of social media data. These techniques are further adopted for social media data, which we will discuss more in the next sections. Finally, we discuss related work and highlight the differences between this work and other surveys in existing literature.

2.1 k-anonymity, l-diversity, and t-closeness

k-anonymity was one of the first techniques introduced for protecting data privacy [171]. The aim of k-anonymity is to anonymize each instance in the dataset so that it is indistinguishable from at least $k-1$ other instances with respect to certain identifying attributes. k-anonymity could be achieved through suppression or generalization of the data instances. The goal here is to anonymize the data such that k-anonymity is preserved for all instances in the dataset with a minimum number of generalizations and suppressions while maximizing the utility of the resultant data. It has been shown that this problem is NP-hard [4]. k-anonymity was initially defined for tabular data, but then researchers start to adopt it for solving privacy issues in social media data. In social media related problems, k-anonymity ensures that users cannot be identified and there are $k-1$ other users with the same set of features that makes these k users indistinguishable. These features may include users' attributes and structural properties.

Although k-anonymity is among the first techniques proposed for protecting the privacy of datasets, it is still vulnerable against specific types of privacy leakage. Machanacajjhala et al. [119] introduces two simple attacks that defeats k-anonymity. The first attack is homogeneity attack, in which the adversary can infer an instance's (in this case, a users in social media) sensitive attributes when sensitive values in an equivalence class lack diversity. In the second attack, the adversary can infer an instance's sensitive attributes when he or she has access to background knowledge even in the case where the data are k-anonymized. The second attack is known as background knowledge attack. Variations of background knowledge attacks are proposed and used for inferring social media users' attributes. The background knowledge could be users' friends' or behavioral information. We will discuss more about different types of the attribute inference attacks problem in Sections 6 and 7.

Table 2. k-anonymity, l-diversity, and t-closeness Applications in User Privacy in Social Media

Technique	Type of Information	Paper
k-degree anonymity	graph structure	[115]
k-neighborhood anonymity	graph structure	[196]
k-automorphism	graph structure	[199]
k-isomorphic	graph structure	[43]
k-anonymity	graph structure and attribute information	[189]
(θ, k)-matching anonymity	graph structure and attribute information	[16]
(k, d)-anonimity	graph structure and attribute information	[17]
l-diversity	attribute information	[119]
t-closeness	attribute information	[107]

To protect data against homogeneity and background knowledge attacks, Machanacajjhala et al. [119] introduce the concept of l-diversity. It ensures that the sensitive attribute values in each equivalence class are diverse. More formally, a set of records in an equivalence is l-diverse if the class contains at least l *well represented* values for the sensitive attributes. The dataset is then l-diverse if every class is l-diverse. Two instantiations of the l-diversity concept are then introduced, entropy l-diversity and recursive (c, l)-diversity. With entropy l-diversity, each equivalence must not only have enough different sensitive values, but also each sensitive value must be distributed evenly enough. More formally, the entropy of the distribution of sensitive values in each equivalence class is at least $log(l)$. For recursive (c, l)-diversity, the most frequent value should appear frequent enough in the dataset. Interested readers could refer to the work of Reference [119] for more details.

After l-diversity, Li et al. [107] studies the vulnerabilities of l-diversity and introduce a new privacy concept, t-closeness. They show that l-diversity cannot protect the privacy of data when the distribution of sensitive attributes in the equivalence class is different from the distribution in the whole dataset. If the distribution of sensitive attributes is skewed, then l-diversity presents a serious privacy risk. This attack is known as the skewness attack. l-diversity is also vulnerable against similarity attacks. This attack can happen when the sensitive attributes in an equivalence class are distinct but semantically similar [107]. Li et al. [107] thus introduce a new privacy concept, t-closeness, which ensures that the distribution of a sensitive attribute in any equivalence class is close to the distribution in the overall table. More formally speaking, an equivalence class satisfies t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution in the whole dataset is no more than a certain threshold. The whole dataset is said to have t-closeness if all equivalence classes have t-closeness. It is valuable to mention that t-closeness protects the data against attribute disclosure but not identity disclosure.

k-anonymity, l-diversity, and t-closeness are further adopted for unstructured social media data. Table 2 summarizes different approaches that leverage adopted versions of these techniques for privacy problems in social media. These works are discussed more in the following sections.

2.2 Differential Privacy

Differential privacy is a powerful technique that protects a user's privacy during statistical query over a database by minimizing the chance of privacy leakage while maximizing the accuracy of queries. It is introduced by Dwork et al. [52, 53] and provides a strong privacy guarantee. The intuition behind differential privacy is that the risk of user's privacy leakage should not be increased as

a result of participating in a database [52]. In particular, it imposes a guarantee on the data release mechanism rather than the dataset itself. The privacy risk is also evaluated according to the existence or absence of an instance in the database. Differential privacy assumes that data instances are independent from each other and guarantees that existence of an instance in the database does not pose a threat to its privacy as the statistical information of data would not change significantly in comparison to the case that the instance is absent [52, 53]. This way, the adversary cannot infer whether an instance is in the database or not or which record is associated with it [92].

Definition 2.3 (Differential Privacy). Given a query function $f(.)$, a mechanism $K(.)$ with an output range $\mathcal{R}$ satisfies ϵ-differential privacy for all datasets $\mathcal{D}_1$ and $\mathcal{D}_2$ differing in at most one element *iff*:

$$\frac{Pr[K(f(\mathcal{D}_1)) = R \in \mathcal{R}]}{Pr[K(f(\mathcal{D}_2)) = R \in \mathcal{R}]} \leq e^{\epsilon}. \tag{1}$$

Here ϵ is called privacy budget and large values of ϵ (e.g., 10) results in large e^{ϵ} and indicates that large output difference could be tolerated and hence we have large privacy loss. This is because the adversary can infer the change in the database according to the large change of the query function $f(.)$. On the other hand, small values of ϵ (e.g., 0.1) indicate that small privacy loss could be tolerated. Query function $f(.)$ can be thought of as a request about value of a random variable and mechanism $K(.)$ is also a randomized function that can be considered as an algorithm that returns the results for the query function, possibly with some noise. To make it more clear, let us assume that we have a dataset containing every patient information. An example of the query function $f(.)$ could be the question, *How many people have the disease x?* The mechanism $K(.)$ could be any algorithm that finds the answer to this question. The output range $\mathcal{R}$ for mechanism $K(.)$ in this example is $\mathcal{R} = \{1, 2, \ldots, n\}$, where n is the total number of patients in the dataset.

Differential privacy models could be either interactive or non-interactive. Assume that the data consumer executes a number of statistical queries on the same dataset. In the interactive models, the data publisher responds to the customer with $K(f(\mathcal{D}))$, where $K(.)$ perturbs the query results to achieve the privacy guarantees. In non-interactive models, the data publisher designs a mechanism $K(.)$, which transforms the original data $\mathcal{D}$ into a new anonymized dataset $\mathcal{D}' = K(f(\mathcal{D}))$. The perturbed data $\mathcal{D}'$ are then returned to the consumer, which is ready for arbitrary statistical queries.

A common way of achieving differential privacy is through adding random noises, i.e., Laplacian or Exponential to the query answers [52]. The Laplacian mechanism is a popular technique for providing ϵ-differential privacy that adds Laplace noise drawn from Laplace distribution. Since ϵ-differential privacy is defined over the query function and holds for all datasets according to Equation (1), the amount of added noise only depends on the sensitivity of the query function. Sensitivity of the query function is further defined as:

$$\Delta(f) = \max \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_1 \tag{2}$$

for any $\mathcal{D}_1$ and $\mathcal{D}_2$ that differ in at most one element. $\|.\|_1$ denotes the l_1 norm.

The added Laplacian noise is then drawn from $Lap(\Delta(f)/\epsilon) \propto e^{-\epsilon/\Delta(f)}$, and the output result considering differential privacy constraint will be $K(f(\mathcal{D})) = f(\mathcal{D}) + Y$, where $Y \sim Lap(\Delta(f)/\epsilon)$. The mechanism $K(.)$ works best when $\Delta(f)$ is small as it introduces the least noise. The larger the sensitivity of a query, the less privacy risks can be tolerated, as removing any instance from the dataset would change the output of the query more. Note that the sensitivity basically captures how a great difference (between the value of $f(.)$ on two datasets differing in a single element) must be hidden by the additive noise generated by the data publisher. Note that recent studies show that the dependency between instances in the dataset will hurt the differential privacy guarantees [92, 113].

Table 3. Differential Privacy Applications in User Privacy in Social Media

Type of Information	Reference
graph structure	[113, 152, 162, 179, 182]
recommender systems	[66, 76, 89, 120, 128, 130, 166, 197, 198]
textual data	[191]

There also exists a relaxed version of ϵ-differential privacy, known as (ϵ, δ)-differential privacy, which was developed to deal with very unlikely outputs of $K(.)$ [52, 53]. It could be defined as:

Definition 2.4 ((ϵ, δ)-differential privacy). Given a query function $f(.)$, a mechanism $K(.)$ with an output range $\mathcal{R}$ satisfies (ϵ, δ)-differential privacy for all datasets $\mathcal{D}_1$ and $\mathcal{D}_2$ differing in at most one element *iff*:

$$Pr[K(f(\mathcal{D}_1)) = R \in \mathcal{R}] \leq e^{\epsilon} \times Pr[K(f(\mathcal{D}_2)) = R \in \mathcal{R}]. \quad (3)$$

Table 3 summarizes different works that utilize differential privacy in social media data. All these works are discussed more later.

2.3 Related Work

There are multiple relevant surveys related to the privacy of data and privacy-preserving approaches [1, 5, 54, 59, 82, 86, 159, 165, 176, 193]. Fung et al. [59] reviews privacy-preserving data publishing methods for relational data such as k-anonymity, l-diversity, t-closeness and their other variations. These methods are compared in terms of privacy models, anonymization algorithms, and information metrics. Zhelva et al. [193] review the concepts of privacy issues in tabular data and introduce new privacy risks in graph data. Multiple surveys focus on reviewing graph data privacy risks [1, 82, 86, 165]. Sharma et al. [165] are among the first works that reviews k-anonymity and randomization-based techniques for anonymizing graph data. Another overview by Abawajy et al. [1] presents the threat model for graph data and classified the background knowledge that is used by adversaries to breach the privacy of users. They also review and classify state-of-the-art approaches for anonymizing graph data. Ji et al. [82, 86] conducted a survey on graph data anonymization, de-anonymization attacks, and de-anonymizability quantification. Another way of sanitizing data is by providing algorithms that are provably privacy-preserving and ensure no sensitive information leak from the data [193]. There is a thorough survey [176] on privacy-preserving data mining, which studies different privacy-preserving data mining approaches. Another work, from Agrawal et al. [5], proposes algorithms to perturb data values by adding random noise to them. Another set of works focuses on developing privacy-preserving association mining rules to minimize privacy loss [54, 159].

In this work, we go one step further and review all aspects of social media data that could lead to privacy leakage. Social media data are highly unstructured and noisy and inherently different from relational and tabular data. Therefore, other approaches are designed specifically to study privacy risks in the context of user-generated data in social media platforms. Different from previous works, we not only reviews state-of-the-art and recent approaches on social graph anonymization and de-anonymization, but we also survey other attribute and identity disclosure attacks that could be performed on the other aspects of user-generated social media data. In addition, we overview and summarize approaches that leaks users' profile attribute and location information utilizing their other online activities. We also survey author identification techniques that incorporate various pieces of user-generated information such as user profiles and textual posts to re-identify users. Besides, we cover more recent works related to privacy leakage in social media that are not

covered in the work of Zhelva et al. [193]. Furthermore, we include many new techniques related to the privacy of social graphs that are not included in previous surveys [1, 82, 86, 165].

In summary, to the best of our knowledge, this is the first and most comprehensive work that systematically surveys and analyzes the advances of research on privacy issues in social media.

3 SOCIAL GRAPHS AND PRIVACY

A large amount of data generated by users in social media platforms has graph structure. Friendship and following/followee relations, mobility traces (e.g., WiFi contacts, Instant Message contacts), and spatio-temporal data (latitude, longitude, timestamps) all could be modeled as graphs. This mandates paying attention to privacy issues of graph data. We will first overview graph de-anonymization works and then survey the proposed solutions for anonymizing graph data.

3.1 Graph De-anonymization

The work of Backstrom et al. [19] was among the first that studied the privacy breach problem related to the social network's graph structure. These attacks could be categorized as either a seed-based or seed-free approach according to whether pre-annotated seed users existed or not. Seed users are those whom their identity are clear for the attacker. Backstrom et al. [19] is among the first seed-based approaches. This work introduces both active and passive attacks on anonymized social networks. In active attacks, the adversary creates k new user accounts (a.k.a. Sybils) and links them to the set of predefined target nodes before the anonymized graph is produced. Then it links these new accounts together to create a subgraph H. After publishing the anonymized graph, the attacher looks for the subgraph H and then locates and re-identifies targeted nodes in the published graph. The main challenge here is that the subgraph H should be unique enough to be found efficiently. In passive attacks, the adversary is an internal user of the system and no new account is created. The attacker then de-anonymizes the users connected to him after the graph data is released. This attack is susceptible to Sybil defense approaches [8] and wrongly assumes that attackers can always change the network before its release.

Another work, from Narayanan et al. [140], introduces an improved attack that does not need compromised accounts or Sybil users to perform the attack. This work assumes that the attacker has access to a different network whose membership has overlap with the original anonymized network. This auxiliary graph is also known as background or auxilary graph knowledge. It also assumes that the attacker has the information of a small set of users, i.e., seed users, who are present in both networks. Narayanan et al. [140] discuss different ways of collecting background knowledge. For example, if the attacker is a friend of a portion of the targeted users, then he or she knows all the details about them [98, 170]. Another approach is paying a set of users to reveal information about themselves and their friends [106]. Crawling data via social media API or using compromised accounts as discussed in active attack are other approaches for gathering background knowledge. Social graph de-anonymization attack in social media could be then formally defined as:

Definition 3.1 (Social Graph De-anonymization Attack [57, 140]). Given an auxiliary/ background graph $G_1 = (V_1, E_1)$ and a target anonymized graph $G_2 = (V_2, E_2)$, the goal of de-anonymization is to find identity disclosures in the form of $1-1$ mappings as many and accurately as possible. An identity disclosure indicates that the two nodes $i \in V_1$ and $j \in V_2$ actually correspond to the same user.

3.1.1 Seed-based De-anonymization. Seed-based de-anonymziation approaches have two main steps. In the first step, a set of seed users are mapped from the anonymized graph to the background/auxiliary graph knowledge and thus are re-identified. In the second step, the mapping and de-anonymization is propagated from the seed users to the other remaining unidentified

users. Similarly, the work of Narayanan et al. [140] starts from re-identifying seed users in an anonymized and auxiliary graph. Then, other users are re-identified by propagating mappings based on seed users pairs. Structural information such as user's degree, user's eccentricity, and edge directionality are used to heuristically measures the strength of match between users. A straightforward application of this de-anonymization attack with less heuristics is predicting links between users [138].

Yartseva et al. [185] propose a percolation-based de-anonymization approach that maps every pair of users in both graphs (background knowledge and anonymized graphs) that have more than k neighboring mapped pairs. The only parameter of this work is k, which is a predefined mapping threshold and does not require a minimum number of users in the seed set. Another similar work, from Korula et al. [99], proposes a parallelizable percolation-based attack with provable guarantees. It again starts with a set of seed users who are previously mapped and then propagates the mapping to the remaining network. Two users will be mapped if they have a specific number of mapped neighbors. Their approach is robust to malicious users and fake social relationships in the network.

In another work, Nilizadeh et al. [142] propose a community-based de-anonymization attack using the idea of divide-and-conquer. Community detection has been extensively studied in the literature of social network analysis [12, 184] and has been used in variety of tasks such as trust prediction [24] and guild membership prediction [13, 69]. In this work, the attacker first partitions both graphs (i.e., anonymized and knowledge graphs) into multiple communities. It then maps communities by creating a network of communities in both graphs. Then users within mapped communities are re-identified and matched together. Mappings are then propagated to re-identify the remaining users. This attack uses similar heuristics as [140] to measure the mapping strength between users.

Ji et al. [80, 81] study de-anonymizability of social media graph data based on seed-based approaches under both the Erdos-Renyi and a statistical model. Similarly to Reference [83], they specified the structure conditions for both perfect and partial de-anonymization. Chiasserini et al. [45, 55] also study the problem of user de-anonymization according to their structural information under the scale-free user relation model. This assumption is more realistic, since users degree-distribution in social media follows power-law distribution, a.k.a. scale-free. Their results show that the information of a large portion of users in the seed set is useless in re-identifying users. This because of the large inhomogeneities in the users degree. This suggests that given a network with n users, the order of $n^{\frac{1}{2}+\epsilon}$ (for any arbitrarily small ϵ) seeds are needed to successfully de-anonymize all users when seeds are uniformly distributed among the vertices. Chiasserini et al. [45, 46] also propose a two-phased percolation graph matching-based attack similar to that in Reference [185].

Bringmann et al. [38] also propose an approach that uses n^{ϵ} seed nodes (for an arbitrarily small ϵ) for a graph with n nodes. This is an improvement over the state-of-the-art structure-based de-anonymization techniques that need $\Theta(n)$ seeds [99]. This approach then finds a signature set for each node as the intersection of its neighbors and previously re-identified nodes. It then defines criterion that further is used to decide if two signatures originate from same nodes with high probability or not, i.e., if the similarity of two nodes signature is more than n^c ($c > 0$ is a constant), then the two nodes are mapped together. Local sensitivity hashing technique [78] is also used to reduce the number of comparisons needed for the de-anonymization attack. Theoretical and empirical analysis of their work show that the attack is performed in quasilinear time.

Manasa et al. [150] propose another seed-based attack against anonymized social graphs that has two steps. In the first step, it identifies a seed sub-graph of users with known identities. As discussed earlier in Reference [19], this sub-graph could be injected by an attacker or it could

Table 4. Summary and Comparison of Seed-based Methods for Graph De-anonymization

Technique \ Properties	With theoretical guarantees	Withouttheoretical guarantees	No limitation on number of seeds
Percolation based	[45, 46, 99]	[185]	[45, 46]
Clustering based	[45, 46, 142]	-	-
Seed extension based	[38, 80, 81, 140]	-	[150]

even be a small group of users that the attack is able to re-identify. In the second step, it extends the seed set based on the users' social relations and re-identifies the remaining users. For each mapping iteration, the algorithm re-examines previous mapping decisions, given new evidence regarding re-identified nodes. This attack does not have any limitation on the size of the initial seed and the number of links between seeds. Another recent work, by Chiasserini et al. [46], incorporates clustering for de-anonymization attacks. Their attack uses various levels of clustering and their theoretical results highlight that clustering can potentially reduce the number of seeds in percolation-based de-anonymization attacks due to its wave-like propagation effect. This attack is a modified version of that in Reference [185], which starts from a small set of seed users and then expands seed set to the closest neighbors of the users in the seed set and repeat the re-identification procedure. In this version, two users are mapped if they have a sufficiently large number of neighbors among the mapped pairs.

To sum-up seed-based graph de-anonymization techniques could be categorized into three groups, percolation-based, clustering-based and seed extension-based works. Table 4 summarizes existing works according to the utilized technique and their properties.

3.1.2 Seed-free De-anonymizatoin. The efficiency of most of seed-based approaches depends on the size of seed set. Seed-free de-anonymization attacks have been developed to solve his issue. Pedarsani et al. [149] present a Bayesian model that starts from the users with the highest degree and iteratively solves a maximum weighted bipartite graph matching problem. This algorithm iteratively updates fingerprints of all users. The goal in the maximum bipartite graph matching problem is to find a maximum matching between two parties so that each vertex is the endpoint of exactly one of the chosen edges.

Moreover, Ji et al. [83, 84] propose to use optimization-based methods to minimize an error function iteratively. More specifically, in each iteration of this attack, two candidate sets of users are selected from the anonymized and background graphs. Then users in the set from the anonymized graph are mapped (de-anonymized) to users in background graph by minimizing an error function defined by the edge difference caused by a mapping scheme. In particular, Ji et al. [83] quantify the structure-based de-anonymization under the Configuration model [141] and drive structural conditions for perfect and partial de-anonymization. The configuration Model generates a random graph given a degree sequence by randomly assigning edges to match the given degree sequence [141].

Another recently developed group of techniques leverages additional sources of information besides structural network to re-identify social media users in anonymized data. This information includes user interactions (e.g., commenting, tweeting) or non-personal identifiable information that is associated with users and is shared publicly such as gender, education, country and interests [64]. This combination of structural and exogenous sources of information could increase the of risk user privacy. Zhang et al. [190] study the privacy breach problem in anonymized heterogeneous networks. They first introduce a privacy risk measure based on the potential loss of the user and the number of users who have same value. They then propose a de-anonymization

algorithm that incorporates the defined privacy risk measure. For each target user, this framework first finds a set of candidates based on entity attribute matches in the heterogeneous network and then narrows down this candidate set by comparing the neighbors (which are found via heterogeneous links) of the target user and each candidate.

Fu et al. [56, 57] propose to use structural and descriptive information. Descriptive information is defined as attribute information such as name, gender, and birth year. This work first proposes a new definition of user similarity, i.e., two users are similar if their neighbors match to each other as well. However, similarity of neighbors also depends on the similarity of users. Therefore, Fu et al. model similarity as a recursive problem and solves it iteratively. Then, they reduce the de-anonymization problem to a complete weighted bipartite graph matching that is solved with Hungarian algorithm [101]. These weights here are calculated based on the users similarities.

In another work, the effect of user attribute information as an exogenous source of information on de-anonymizing social networks is studied [154]. In particular, this work incorporates semantic background knowledge of adversary in the de-anonymization process and models it using knowledge graphs [79]. This approach simultaneously de-anonymizes and infers users attributes (we will discuss user profile attribute inference attack later in Section 5). The adversary first models both the de-anonymized dataset and the background knowledge as two knowledge graphs. Then, she makes a complete adversary weighted bipartite graph. Each weight indicates the structural and attribute similarity between corresponding nodes in the anonymized and knowledge graphs. The de-anonymization problem will be then reduced to a maximum weighted bipartite matching problem that can be furthered reduced to a minimum cost maximum flow problem. Attacker prior semantic knowledge could be obtained via different ways such as common sense, statistical information, personal information, and network structural information.

Ji et al. [87] also study the same problem and show theoretically and empirically that using attribute information alongside structural information could result in a great privacy loss even in an anonymized dataset in comparison to the cases where the data only consists of structural information. They further propose the De-SAG de-anonymization framework, which incorporates both attribute and structural information. It first augments both types of information into a structure-attribute graph. De-SAG has two variants, i.e., user based and set based. In user-based De-SAG, the proposed de-anonymization approach first selects the most similar candidates to the target user from background/auxiliary knowledge graph based on similarity of their attributes. Next, the target user will be mapped to one of the selected candidates based on their structural similarity. In set-based De-SAG, for each iteration, two sets of users are selected from anonymized graph and knowledge graph, respectively. Then, the de-anonymization problem reduces to a maximum weighted bipartite graph Matching problem and users in these two sets are mapped to each other using Hungarian algorithm [101]. Note that the similarity of users are again calculated according to their attribute and structural information.

In another work, by Lee et al. [105], a blind de-anonymization technique is proposed in which the adversary does not need to have any background information. Inspired by the idea of dK-series for chatacterizing structural characteristics of a graph, they propose nK-series to describe structural features of each user by exploiting his multi-hop neighbors information. In particular, nKi captures the degree histogram of the user's i-hop neighbors. Then, a structure score is calculated for each user (in both the anonymized graph and the background knowledge graph) based on his diversity score (calculated according to nK-series scores) and his relationships with all other non-reidentified users in the network. It then uses this information to re-identify all users in the anonymized social graph by leveraging pseudo relevance feedback support vector machines. Backes et al. [18] develop an attack that infers social links between users based on their mobility profiles without using any additional information about existing relations between users. Their

Table 5. Summary and Comparison of Seed-free Methods for Graph De-anonymization

Technique \ Properties	Homogeneous networks	Heterogeneous networks
Maximum weighted bipartite matching based	[149]	[56, 57, 87, 154]
Error minimization based	[83]	[28, 190]
Learning based	[105, 163, 164]	[18]

approach first constructs mobility profile for each user by obtaining random walk traces from the user-location bipartite graph and using skip-gram [131] to obtain features in a continuous vector space. It then infers the links based on the similarity of their mobility profile.

Beigi et al. [26, 28] introduce a new adversarial attack that does not need to have any background information before initiating the attack. This attack is designed for heterogeneous social media data, which consists of different aspects (i.e., textual and structural) and shows that anonymizing all aspects of data is not sufficient. This attack first extracts the most revealing information for each user in the anonymized dataset and then accordingly finds a set of candidate users. Each user is finally mapped to the most probable candidate user. Sharad et al. [164] propose to formulate the problem of graph de-anonymization in social networks as a learning task. They use one-hop and two-hop neighborhood degree distributions to represent each user. The intuition behind this selection is that two nodes refer to the same user if their neighborhoods also matches to each other. These features are further used to train a classifier to learn the degree deviation for identical and non-identical user pairs. In another work, Sharad et al. [163] go even further and propose a new generation of de-anonymization attacks that is heuristic free, seedless, and considered a learning problem. They use the same set of structural features as proposed in Reference [164] and then de-anonymize the sanitized graph by re-identifying users with high degree first and then use them to attack low-degree nodes. Mappings are then frozen and propagated to the remaining nodes to discover new set of mappings.

Table 5 categorizes reviewed works based on the used technique and the fact that if they are applicable on heterogeneous or homogeneous graph networks.

3.1.3 Theoretical Analysis and De-anonymization. Another set of works studies de-anonymization attacks from the theoretical perspective of view. For example, Liu et al. [113] theoretically study the vulnerability of differential privacy mechanisms against de-anonymization attacks. Differential privacy provides protection against even the strongest attacks in which the adversary knows the entire dataset except one entry. However, differential privacy assumes the independence between dataset entities that is not correct in most real-world applications. This work introduces a new attack in which the probabilistic dependence between dataset entries are calculated and then leveraged to infer users' sensitive information from differentially private queries. The attack is also tested on graph data in which users' degree distributions is published differentially privately.

Lee et al. [104] also study the theoretical quantification for relating the anonymized graph data vulnerability against de-anonymization attacks. In particular, they study the relation between application specific anonymized data utility (i.e., quality of data) and capability of de-anonymization attacks. They define local neighborhood utility and global structure utility. They theoretically show that under certain conditions for each of defined utilities, the probability of successful de-anonymization approaches one with the increase of number of users in data. Their foundations could be used to evaluate the effectiveness of the de-anonymization/anonymization techniques.

Recent research by Fu et al. [58] studies the conditions under which the adversary can perfectly de-anonymize user identities in social graphs. In particular, they theoretically study the cost of quantifying the quality of the mappings. Community structures are also parameterized and leveraged as side information for de-anonymization. They study two different cases in which the community information is available for both background knowledge and anonymized graphs or only for one of them. They showed that perfectly de-anonymizing graph data with community information in polynomial time is NP-hard. They further propose two algorithms with approximation guarantees and lower time complexity by relaxing the original optimization problem. The main drawback of this study is the assumption of disjoint communities, which fails to reflect the real-world situations. Wu et al. [181] extend Fu et al.'s study by considering overlapping communities. In contrast to Fu et al.'s work [58], which uses Maximum a Posteriori estimation to find the correct mappings, Wu et al. introduces a new cost function, Minimum Mean Square Error, which minimizes the expected number of mismatched users by incorporating all possible true mappings.

There are different surveys [1, 82, 86, 104] on quantification and analysis of graph de-anomyziation techniques that study a portion of covered works here in terms of scalability, robustness, and practicability. Interested readers can refer to these surveys for further readings.

3.2 Graph Anonymization

Another research direction in protecting privacy of users in graph data is studying graph anonymization techniques. Existing anonymization approaches use different techniques and mechanisms and could be categorized mainly into five categories: k-anonymity-based approaches [43, 115, 189, 196, 199], edge manipulation techniques [188], cluster-based techniques [31, 70, 114, 134, 174], random walk-based techniques [116, 134], and differential privacy-based techniques [152, 162, 179, 182]. We discuss each of these categories later.

3.2.1 K-anonymity-based Approaches. The aim of k-anonymity methods is to anonymize each user/node in the graph so that it is indistinguishable from at least $k-1$ other users [171]. Liu et al. [115] proposed an anonymization framework for k-degree anonymization in which for each user, there are at least k other users with the same degree. The goal of this approach to add/delete the minimum number of edges to preserve k-degree anonymity. This algorithm has two steps. In the first step, given the degree sequence of the original graph, a k-degree anonymized version of the degree sequence is constructed and then in the second step, the anonymized graph is built based on the anonymized degree sequence. In another work [196], Zhou et al. aim to achieve k-neighborhood anonymity. They consider the assumption that the adversary knows the subgraph constructed by the immediate neighbors of a target node. In the first step of the anonymization, one-hop neighborhoods of all users are extracted and encoded in a way that isomorphic neighborhoods could be easily identified. In the second step, users with similar/isomorphic neighborhoods are grouped together until size of each group is at least k. Then, each group is anonymized satisfying k-neighborhood anonymity as each neighborhood has at least $k-1$ isomorphic neighborhoods in the same group. Eventually, this approach anonymizes the graph against neighborhood attacks.

Zou et al. [199] propose a k-automorphism-based framework that protects the graph against multiple attacks including the neighborhood attack [196], degree-based attack [115], hub-fingerprint attack [70], and subgraph attack [70]. A graph is k-authomorphic if there exists $k-1$ automorphic functions in the graph and for each user in the graph, the attacker cannot distinguish it from her $k-1$ symmetric vertices. The proposed approach first partitions the graph into n blocks and then clusters blocks into m groups (graph partitioning step). In the second step, alignments of blocks are obtained and original blocks are replaced with alignment blocks (block alignment step). In the last step, edge copy is performed to get the anonymized graph. Edge copy adds $k-1$

edges between $k-1$ pairs $(F_a(u), F_a(v)))$ $(a = 1, 2, \ldots k-1)$, where $F_a(.)$ is the automorphic function and u and v are users in the social graph. Authors also propose the use of generalized vertex ID's for handling dynamic data releases. Another similar work, by Cheng et al. [43], proposes a k-isomorphism anonymization approach. A graph is k-isomorphic if it is consisted of k disjoint subgraphs and all subgraphs pairs are isomorphic. In the first step, the graph is partitioned into k subgraphs with the same number of vertices. Then, edges are added or deleted so that these subgraphs are isomorphic. This approach protects the published graph against neighborhood attacks [196].

Yuan et al. [189] incorporate semantic and graph information together to achieve personalized privacy anonymization. In particular, they consider three different levels for attacker's knowledge regarding the target user, (1) only attribute information, (2) both attribute and degree information, and (3) combination of attribute, node degree, and neighborhood's information. They accordingly propose three levels of protection to achieve k-anonymity. For level 1 protection, their approach considers label generalization. For the level 2 anonymization, it uses node/edge adding approach as well. For the level 3 protection, it uses edge label generalization.

3.2.2 Edge Manipulation-based Approaches. Edge manipulation and randomization algorithms for social graphs usually utilizes edge-based randomization strategies to anonymize data such as random edge adding/deleting and random edge switching [188]. Ying et al. [188] propose spectrum preserved edge editing that either adds k random edges to the graph and removes another k edge randomly or alternatively switches k edges. In the switching technique, two random edges, (i_1, j_1) and (i_2, j_2), are selected from the original graph edge set E such that $\{(i_1, j_2) \notin E \wedge (i_2, j_1) \notin E\}$. Then edges (i_1, j_1) and (i_2, j_2) are removed, and new edges (i_1, j_2) and (i_2, j_1) are added instead. This method is going to protect the graph against the edge inference attack. Backes et al. [18] also propose a randomization-based approach to preserve the privacy of social links between users in graph data and counteract link inference attacks. In this specific type of attack, the adversary exploits users mobility traces to infer social links between users with the intuition that friends have more similar mobility profiles in comparison to the mobility profiles of two strangers [18]. They utilize three privacy-preserving techniques: hiding, replacement, and generalization of user mobility information. Results show that data publishers need to hide 80% of the location points or replace 50% of them to prevent leakage of information of users social links.

3.2.3 Clustering-based Techniques. Clustering-based approaches group users and edges and only reveal the density and size of the cluster so that individual attributes are protected. Hay et al. [70] propose an aggregation-based method for graph data anonymization that is robust against three types of attacks: neighborhood, subgraph, and hub fingerprint. It models the aggregate network structure by partitioning original graph and describing it at the level of partitions. Partitions are considered as nodes and edges between them makes the edges in the generalized graph. A graph can be then randomly sampled it and be published as the anonymized graph data.

Another cluster-based work [31] proposes two approaches, label list and partitioning, which consider user attributes (i.e., labels) in addition to structural information. In the label list approach, a list of labels are allocated to each user that also includes her true label. This approach first clusters nodes into m classes and then a set of symmetric lists is built deterministically for each class from the set of nodes in the corresponding class. In the partitioning approach, nodes are divided into classes and instead of releasing full edge information, only the number of edges between and within each class is released. This is similar to the generalization approach of Hay et al. [70]. Bhagat et al. also use a set of safety conditions to ensure that the released data do not leak information. The proposed partitioning approach is more robust than the label list technique when facing the

attacks with richer background knowledge. However, the partitioning approach has lower utility than the label list as less information is revealed about the graph structure.

Thompson et al.'s approach [174] protects the graph information against i-hop degree-based attack. They present two clustering algorithms, bounded t-means clustering and union-split clustering. These approaches group users with similar social roles into clusters with a minimum size constraint. Then they utilize the proposed inter-cluster matching anonymization method, which anonymizes the social graph by removing/adding edges according to the users' inter-cluster connectivity. The number of nodes and edges between and within clusters are then released similar to Hay et al.'s approach [70]. Mittal et al. [114] also propose another clustering-based aonymization technique that considers evolutionary dynamics of social graphs such as node/edge addition/deletion and consistently anonymizes the graph. It first dynamically clusters nodes and then perturbed the intra-cluster and inter-cluster links for changed clusters in a way that structural properties of social media graph is preserved. They leverage static perturbation method in Reference [134] to modify intra-cluster links and randomly connect marginal nodes to create fake inter-cluster links according to their degree. The obfuscated graph is robust against the edge inference attack and has higher indistinguishability that is defined from an information theoretic perspective.

3.2.4 Random Walk-based Approaches. Another group of works utilizes random walk idea to anonymize graph data. The idea of random walk has been previously used in many security applications such as Sybil defense [8]. Recent works also use this idea for anonymzing social graphs. The work of Mittal et al. [134] introduces a random-walk-based edge perturbation algorithm. According to this approach, for each node u, a random walk with the length t will be performed starting from one of the u's contacts, v and an edge (u, z) between destination node, z and u will be added with an assigned probability, and the edge (u, v) will be removed accordingly. This probability will decrease as more random walks are performed from u's contacts. Later, Liu et al. [116] improve this approach such that instead of having a fixed length random walk with length t, they utilize a smart adaptive random that its length is learned based on the local structure characteristics. This method first predicts the local mixing timing for each node, which is the minimum random walk length for a starting node to be within a given distance to stationary (distance) node. This mixing time is predicted based on the local structure and limited global knowledge of the graph and is further used to adjust the length of random walk for social graph anonymziation.

3.2.5 Differential Privacy-based Approaches. Recently, many works extend differential privacy [52] to the social graph data. Sala et al. [162] first use dK-series to capture sufficient graph structure at multiple granularities. dK-series is the degree distributions of connected components of size K within a target graph [50, 122]. Then, they partition the statistical representation of the graph captured by dK-series into clusters and use ϵ-differential privacy mechanism to add noise to the representation in each cluster. Another differentially private-based approach [152] scales down the magnitude of added noise by reducing the contributions of challenging records.

In another work, Wang et al. [179] use dK-graph generation models to generate sanitized graphs. In particular, their approach first extracts various information form the original social graph such as degree correlations and then enforces differential privacy on the learned information and, finally, uses perturbed pieces of information to generate an anonymized graph with dK-graph models. Different from the approach in Sala et al. [162], in the specific case of $d = 2$, noise is generated based on the smooth sensitivity rather than global sensitivity. The reason behind this specification is to reduce the magnitude of the added noise. Smooth sensitivity is a smooth upper bound on the local sensitivity when deciding the noise magnitude [143]. Another work, Reference [182], proposes an anonymization approach that satisfies edge ϵ-differential privacy to hide each user's

Table 6. Summary and Comparison of Different Graph Anonymization Methods

Defense \ Attack	Degree based	Neighborhood based	Hub-fingerprint based	Subgraph based	Edge inference	Attribute inference
k-anonymity based	[115, 189, 199]	[43, 189, 196, 199]	[189]	[43, 189]	-	[199]
Edge manipulation	-	[18]	-	-	[188]	-
Cluster based	[174]	[31, 70]	[31, 70]	[31, 70]	[134]	[31]
Random walk based	-	-	-	-	[116, 134]	-
Differential privacy based	[113, 152, 162, 179]	-	-	-	[182]	[113]

Each column refers to the works that are robust againstthe mentioned attack.

connections to other users. They propose to learn how to transform edges to connection probabilities via statistical Hierarchal Random Graphs (HRG) under differential privacy. In particular, their approach infers the HRG by learning the entire HRG model space and sampling an HRG by a Markov Chain Monte Carlo method and generating the sanitized graph according to the sampled HRG while satisfying differential privacy. Their results show that using edge probabilities can result in significant noise scale reduction in comparison to the case where the edges are used directly.

In another work, from Liu et al. [113], it has been shown that differential privacy is not robust to the de-anonymization attacks if there is dependence among dataset entries. Liu et al. [113] also propose a stronger privacy notion, dependent differential privacy in which it incorporates the probabilistic dependence between the tuples in a statistical database. They then propose an effective perturbation framework that provides privacy guarantees. Their result show that more noise should be added when there is dependency between tuples. The added noise is also dependent on the sensitivity of two tuples as well as the dependence relationship between them. They evaluate their proposed framework on graph data to sanitize the degree distribution of the given graph.

Ji et al. [82, 86] and Abajaway et al. [1] study the defense and attacking performance of a portion of existing social graph anonymization and de-anonymization techniques. Ji et al. [82, 86] have also performed a thorough theoretical and empirical analysis on a portion of existing related papers. Results demonstrate that anonymized social graphs are vulnerable to de-anonymization attacks.

To sum up, Table 6 categorizes reviewed works with respect to the utilized technique, i.e., k-anonymity, edge manipulation, cluster based, random walk based, and differential privacy based. Each column in Table 6 refers to the type of graph de-anonymization attack and correspondingly the works that are robust against the mentioned attack.

4 AUTHORS IN SOCIAL MEDIA AND PRIVACY

People have the right to have anonymous free speech over different topics such as politics. However, an author's identity can be unmasked by adversaries through providing her real name or IP address to a service provider. However, authors can use tools such as Tor to protect their identity at the network level. Manually generated content will always reflect some characteristics of the person who authored it. For example, some anonymous online author is prone to several specific spelling errors or has other recognizable idiosyncrasies [137]. These characteristics could be enough to figure out whether authors of two pieces of content are same or not. Therefore, with material authored by the true identity of the author, the adversary can discover the identity of a content posted online by the same author anonymously. Identifying the author of a text according to her writing style, a.k.a. stylometry, has been studied for a long time [135, 169]. With the adverse

of machine learning techniques, researches start to extract textual features and discriminate between 100 and 300 authors [2]. The application of author identification includes identifying authors of terroristic threats and harassing messages [42], detecting fraud [3], and extracting demographic information [95].

Privacy implications of stylometry have been studied recently. For example, Rao et al. [156] investigate whether people who are posting under different pseudonyms to USENET newsgroup can be linked based on their writing style. They use a dataset of 117 people having 185 different pseudonyms and exploit function words and Principal Component Analysis (PCA) to perform matching between newsgroups posting and email domains. Another work, from Koppel et al. [96, 97], studies author identification at the scale of over 10,000 blog authors. They use 4-grams of characters, which is a context specific feature. The problem with this work is that it is not clear whether their approach is solving author recognition or context recognition. In another work, Koppel et al. [95] use both content-based and stylistic features to identify 10,000 authors in the blog corpus dataset. There are also several works on identifying authors of academic papers under blind review based on the citations of the paper [37, 73] or other sources from unblind texts of potential authors [136].

Narayanan et al. [137] propose another author identification attack that exploits 1,188 real-valued features from each post, such as frequency of characters, capitalization of words, syntactic structure (extracted by Stanford Parser [93], e.g., noun phrases containing a personal pronoun, noun phrases containing a singular proper noun), and distribution of word length. These features capture the writing style of the author regardless of the topic at hand and can re-identify large number of authors. However, this approach will not work when authors anonymize their writing style. Almishari et al. [10] proposed a new linkage attack that investigates the linkability of prolific reviews that users post on social media platforms. More specifically, given a subset of information on reviews made by an anonymous user, this approach seeks to map it to a known identified record. This approach first extracts four types of tokens: (i) unigrams, (ii) digrams, (iii) ratings, and (iv) category of reviewed entity. Then, it uses Naive Bayes and Kullback–Leibler (KL) divergence models to re-identify the anonymized information. This approach could be also used for identity disclosure attack across multiple platforms using people's posts and reviews.

Bowers et al. [36] propose an anonymization approach that uses iterative language translation to conceal one's writing style. This approach first translates English text into another foreign language (e.g., Spanish, Chinese, etc.) and then turns it back to English again for three iterations. Another work, from Nathan et al. [121], evaluates Bowers's work by introducing a feature selection approach, namely Generative and Evolutionary Feature Selection (GEFES), over the set of predefined features that mask out non-salient previously extracted features. Both Reference [36] and Reference [121] are tested over a set of blog posts by users and the results show the efficiency of ILT-based anonymization. A recent work is also proposed by Zhang et al. [191] that anonymizes users' textual information before publishing user-generated data. This approach first introduces a verified version of differential privacy specified for textual data, namely, ϵ-Text Indistinguishability, to overcome the curse of dimensionality problem when original differential privacy is deployed on high-dimensional textual data. It then proposes a framework that perturbs user-keyword matrix by adding Laplacian noise to satisfy ϵ-Text Indistinguishability. Results confirms both the utility and privacy of the data.

5 SOCIAL MEDIA PROFILE ATTRIBUTES AND PRIVACY

A user's profile includes her self-disclosed demographic attributes such as age, gender, majors, cities she loved, and so on. To address the privacy of users, social networks usually offer the option for users to limit the access to their attributes, i.e., they are only visible to friends or friends of

friends. A user could also create a profile without explicitly disclosing any attribute information. A social network thus is a mixture of both private and public user information. However, there exists one privacy attack that focuses on inferring users' attributes. This attack is known as attribute inference attack and it leverages publicly available information of users in social networks to infer missing or incomplete attribute information [63].

The attacker could be any party who is interested in this information such as social network service providers, cyber criminals, data brokers, and advertisers. Data brokers benefit from selling individuals' information to other parties such as banks, advertisers, and insurance companies.[1] Social network providers and advertisers leverage users' attribute information to provide more targeted services and advertisements. Cyber criminals exploit attribute information to perform targeted social engineering, spear phishing,[2] and backup authentication attacks [68]. This attribute information could be also used for linking users across multiple sites [62] and records (e.g., vote registration records) [132, 171]. Existing attacks could be categorized into three groups: friend based, behavior based, and friend and behavior based.

5.1 Friend-based Profile Attribute Inference

Friend-based approaches use homophily theory [127], which states that two friends are more probable to share similar attributes rather than two strangers. Following this intuition, if most of a user's friends study at Arizona State University, then she is more likely studying in the same university. He et al. [71] first constructs a Bayesian network from a user's social neighbors and then uses it to model the causal relations among people in the network and thus obtains the probability that the user has a specific attribute. The main challenge in this approach is its scalability as Bayesian inference is not scalable to the millions of users in social networks. Another work, by Lindamood et al. [111], uses Naive Bayes classification algorithm to infer a user's attributes by exploiting features from her node trait (i.e., other available attributes information) and link structures (i.e., friends). However, this approach is not usable for a user who does not share any attributes. In the other work, Reference [173], the authors propose an approach that leverages friends' activities and information to infer a user's attributes. These features from friends and wall posts are then exploited into a multi-label classifier. The authors then propose a multi-party privacy approach that defends against attribute inference attacks. This approach enforces mutual privacy requirements for all users to prevent disclosure of users attributes and sensitive information.

Zhelva et al. [192] study how users sensitive attribute information could be leaked through their social relations and group memberships. This friend-based attribute inference attack exploits social links and group information to infer sensitive attributes for each user. Authors propose various algorithms in which it was found LINK was the best among those that only use link information. This method models each user u as a binary vector whose length is the size of the network (i.e., number of users in the network) and the value of each element v is one if u is connected to v. Then, different classifiers are trained over the users with a public profile and then attributes for users with private profiles could be inferred. The GROUP algorithm was the best among the methods that incorporates group information. This method first selects the groups that are relevant to the attribute inference problem using either feature selection approach (i.e., entropy) or manually. Next, relevant groups are considered as features for each node and a classifier model is trained. In the last step, the attributes for targeted users are predicted using the classification model. Mislove et al. introduces a similar approach that leverages users' social links and communities information [133]. Their approach takes some seed users with known attributes as the input and then

[1] https://bit.ly/1AwePQE.
[2] http://www.microsoft.com/protect/yourself/phishing/spear.mspx.

finds the local communities around this seed set using available link information. Then it uses the fact that users in the same community share similar attributes. This approach then infers remaining users' attributes based on the communities they are a member of. The limitation is that this approach is not able to infer attributes for users who are not assigned to any local communities.

Avello et al. [61] propose a semi-supervised profiling approach named McC-Splat. They consider the attribute inference problem as a multiclass classifier. It then learns the attributes' weights according to the user's friends' attributes. Weights here indicate the users' likelihood in belonging to a given attribute value class. Finally, McC-Splat assigns the class with the highest percentile to the target user. The percentile is calculated according to the labeled individuals information. In the other work, from Dey et al. [49], the authors focus on predicting facebook users' ages considering their friendship network information. Although a user's friends list is not fully available for all users, this work uses reverse lookup approach to obtain a partial friend list for each user. Then, they designed an iterative algorithm that estimates users' ages based on friends' ages, friends of friends' ages and so on. They also incorporated other public information in each user's profile such as their high school graduation year to estimate their birth year. Another work, Reference [77], seeks to find a targeted user based on her social network connections and the similarity of attributes between friends. It starts from a source user and continue crawling until it reaches the target user. The navigations are based on the set of target user's known attributes, friendship links between users and their attributes as well. Similarly, Labitzke et al. [102] also study whether profile information of Facebook users could be still leaked through their social relations. A recent work published by Li et al. [110] uses convolutionam neural network (CNN) to infer multi-valued attributes for a target user according to his ego network. A user's ego network is a subset of the original social network based on the user's friends and the social relations among them. CNN can capture the latent relationship between users' attributes and social links.

Another set of works in this category focuses on predicting both network structure (i.e., links) and inferring missing users attribute information [65, 186, 187]. The reason for simultaneously solving these two problems is that users with similar attributes tend to link to one another and individuals who are friends are likely to adopt similar attributes. The work of Yin et al. [186, 187] first creates a social-attribute network graph from an original social graph and user-attributes information, i.e., nodes in the graph are either users or attributes. Edges show the friendship between a pair of users or the relation between a user and attribute. Then, authors use random walk with restart algorithm [175] to calculate link relevance and attribute relevance with regard to a given user. Similarly, Gong et al. [65] transform the attribute inference attack problem to a link prediction problem in the social-attribute network graph. They generalized several supervised and unsupervised link prediction algorithms to predict the links between user-user and user-attributes.

5.2 Behavior-based Profile Attribute Inference

Unlike friend-based approaches, behavior-based inference attacks infer a user's attributes based on the publicly available information regarding her behaviors and public attributes of other users similar to her. Weinsberg et al. [180] propose an approach that infers users' attributes (i.e., gender) according to their behavior toward movies. In particular, each user is modeled with a vector with the size being the number of items. A non-zero value for each vector element demonstrates that the user has rated the item, and zero value means that user has not rated the item. Then, they use different classifiers such as logistic regression, SVM, and Naïve Bayes to infer users' ages. Accordingly, the authors propose a gender obfuscation method that adds movies and corresponding ratings to a given user's profile such that it will be hard to infer the gender of the user while minimally impacting the quality of recommendations the user received. They use three different approaches for movie selection: random, sampled, and greedy strategy. The sampled strategy picks a movie based

on ratings distribution associated with the movies of the opposite gender. The greedy approach also selects a movie with the highest score in the list of movies for opposite gender. Ratings are also added for each movie based on either the average movie rating or the rating predicted using recommendation approaches such as matrix factorization. The greedy movie selection approach with predicted rating has the best results regarding user profile obfuscation. Kosinski et al. [100] follow a similar approach to Reference [180] and construct a feature vector for each user based on Facebook likes. Authors then use logistic regression classifier to infer various attributes for each user.

Another work, from Bhagat et al. [32], proposes an active learning-based attack that infers users' attributes via interactive questions. In particular, their approach involves finding a set of movies and asking users to rate them. Each selection maximizes the confidence of the attacker in inferring users attributes. The work of Reference [41] seeks to infer users attributes based on the different types of music they like. This approach first extracts a user's interests and finds semantic similarity among them. It uses an ontologized version of Wikipeda related to each type of music, exploits topic modeling techniques (i.e., Latent Dirichlet Allocation, LDA [34]), and learns semantic interest topics for each user. Then, a user is predicted to have similar attributes as those who like similar types of musics as the user. In another work, from Luo et al. [117], authors infer household structures of Internet Protocol Television based on the users' watching behavior. Their approach first extracts related features from log-data including TV programs topics and viewing behavior using LDA and low-rank model, respectively. Then, it combines graph-based semi-supervised learning with non-parametric regression and uses it to learn a classifier for inferring the household structure.

5.3 Friend and Behavior–based Profile Attribute Inference

Another category of approaches exploit both social link and user behavior information for inferring users attributes. Gong et al. [63, 64] first make a social-behavior-attribute network (SBA) in which social structures, user behaviors, and user attributes are integrated into a unified framework. Nodes of this graph are users, behaviors, or attributes, and edges represents the relationship between these attributes. Then, they infer a target user's attributes through a vote distribution attack (VIAL) model. VIAL performs a customized random walk from a target user to all other users in the augmented SBA network and assigns probabilities to the users such that a user receives higher probability if it is structurally more similar to the target node in SBA network. The stationary probabilities of attribute nodes are then used to infer attributes of the target user, i.e., the attribute with maximum probability is assigned to the target user. Unlike most of the existing approaches that only use the information of users who have an attribute, a recent work from Ji et al. [88] incorporates information from users who do not have the attribute in the training process as well, i.e., negative training samples. This work associates a binary random variable with each user characterizing whether a user has an attribute or not. Then it learns the prior probability of each user having a specified attribute by incorporating the user's behavior information. Next, it models the joint probability of users as a pairwise Markov Random Field according to their social relationships and uses this model to infer posterior probability of attributes for each target user.

5.4 Exploiting Other Sources of Information for Profile Attribute Inference

These approaches leverage sources of information other than social structures and behaviors, such as writing style [144], posted tweets [9], liked pages [68], purchasing behavior [178], and checked-in locations [195]. A recent research combined identity and attribute disclosure across multiple social network platforms [16]. It defines the concept of (θ, k)-matching anonymity as a measure of identity disclosure risk. Given a user and her identity in a source social network, a matching anonymity set is defined as the set of identities in the target social network with a matching

Table 7. Summary and Comparison of Different Attribute Inference Attacks w.r.t. Utilized Techniques and Leveraged Information in the Attack

Technique \ Information	Friend	Behavior	Friend & Behavior
Community and clustering based	[65, 133, 192]	[41]	-
Random walk based	[186, 187]	-	[63, 64]
Traditional supervised classification models (e.g., SVM, logistic regression, etc.)	[65, 71, 111, 173]	[100, 180]	-
Graphical model based	-	-	[88]
Semi-supervised classification	[61]	[117]	-
Iterative based	[49, 77]	-	-
Active learning	-	[32]	-

probability of more than θ. The user is (θ, k) anonymous if the size of the matching set is k. Another work, by Backes et al. [17], introduces a relative linkability measure that ranks identities within a social media site. In particular, it incorporates the idea of k-anonymity to define (k, d)-anonimity for each user u in social media that captures the largest k subset of identities (including u) who are within a similarity (or dissimilarity) threshold d from u considering their attributes. A recent work from Liu et al. [113] also studies the vulnerability of differential privacy mechanism against the inference attack problem. As stated earlier, differential privacy provides protection against the adversary who knows the entire dataset except one entry. However, differential privacy considers the independence between dataset entities. Liu et al. introduce a new inference attack in which the probabilistic dependence between dataset entries are calculated and then leveraged to infer a user's location information from differentially private queries.

Different from all the works focusing on profile attribute inference, a recent work, Reference [11], brings evasion and poisoning attacks into this problem. This work introduces five variants of evasion and poisoning attacks to interfere with the results of the profile attribute inference:

- **Good/Bad Feature Attack (Evasion)**: The adversary adds good features from one attribute to another while removing bad features from each class to introduce false signals for the predictor.
- **Mimicry Attack (Evasion)**: Adversary samples a set of users from one class and then finds the most similar users in the other class. Good (bad) features are added (removed) for users in the found subsets.
- **Class Altering Attack (Poisoning)**: Adversary randomly chooses users from one class and then flips their class label. This results in higher misclassification rate.
- **Feature Altering Attack (Poisoning)**: The goal is to increase the misclassification rate. She poisons the training data by randomly adding good feature values of one class to another class.
- **Fake Users Addition Attack (Poisoning)**: The attacker poisons the data by removing a set of real users and then injecting fake users into the training dataset.

Table 7 summarizes existing works based on the technique they have used and the type of information leveraged for attribute inference attacks. Utilized techniques could be categorizes into different groups: community and clustering-based, random walk-based, graphical model-based, iterative-based, active learning-based, semi-supervised-based, and traditional supervised methods.

6 SOCIAL MEDIA USERS LOCATION AND PRIVACY

This location disclosure attack is a specific version of attribute inference attack in which the adversary focuses on inferring geo-location information for a given user. The location disclosure attack takes as input some geolocated data and produces some additional knowledge about target users. More precisely, the objective of this attack may be to (1) predict the movement patterns of an individual, (2) learn the semantics of the target user mobility behavior, (3) link records of the same individual, and (4) identify points of interest [60]. Existing works incorporate a given user's friends' known geo-location information [20, 47, 90, 91, 94, 125, 126, 160]. The work of Reference [20] introduces a probabilistic model representing the likelihood of the target user's location based on her friends' location and geographic distance between them. Reference [94] and Reference [126] extend Backstrom et al.'s work [20] and find the target user's friends that are strong predictors of her location.

In another work, Mcgee et al. [125] integrates social tie strength information to capture the uncertainty across multiple location granularities. The reason is that not all relationships in social media are the same and the location of friends with strong ties are more revealing of a user's location. Rout et al. [160] deploy a SVM classifier on a given set of features to predict the target user's location. These features include cities of the target user's friends, number of friends in the same city as the target user and number of reciprocal relationships the target user has per city. Jurgens et al. [90] infer locations by proposing an iterative multi-pass label propagation approach. This approach calculates each target user's location as the geometric median of her friends' locations and it seeks to overcome the sparsity problem when the ground truth data is sparse. The work of Reference [47] extends Reference [90] and limits the propagation of noisy locations by weighting different locations using information such as the number of times the users have interacted.

Another work, from Cheng et al. [44], proposes a probabilistic framework that infers Twitter users' city level location based on the content of their tweets. The idea is that users' tweets include either implicit or explicit location-specific content, e.g., place names, or words or phrases more associated with certain locations (e.g., "howdy" for Texas). It uses lattice-based neighborhood smoothing technique to even out the word probabilities and overcome the tweet sparsity challenge. Hecht et al. [72] also found that only 34% of Twitter users do not provide their real location information or share fake locations or sarcastic comments to fool location inference approaches. They show that a user's location could be inferred using machine learning techniques through the implicit user behavior reflected in their tweets. In another work, Ryoo et al. [161] refine Cheng et al.'s city-level granularity location inference approach [44] to 500-m distance bins. Having GPS-tagged tweets for a set of users, their approach builds geographic distributions of words and computes user location as a weighted center of mass from the user's words. It then uses a probabilistic model and computes the foci and dispersions by binning the distance between GPS coordinates and the word's center by 500 m for computational scalability.

Li et al. [109] introduce a unified discriminative influence model that considers both users' social network and user-centric data (e.g., tweets) to solve the scarce and noisy data challenge for location inference. It first augments social network and user data in a probabilistic framework that is viewed as a heterogeneous graph with users and tweets as nodes and social and tweeting relations as edges. Every node in this graph is then associated with a location and the proposed probabilistic influence model measures how likely an edge is generated between two nodes considering their locations. Another similar work, from Li et al. [108], exploits a user's tweets and social relations to build a complete location profile that infers a set of multiple long-term geographic location scopes related to her, which not only includes her home location, but also other related ones, e.g., work space. Their approach captures the user's friends' locations as well.

Srivatsa et al. [168] propose a de-anonymization attack that exploits a user's friendship information in social media to de-anonymize users mobility traces. The idea behind this approach is that people meet those who have a relationship with them and thus they could be identified by their social relationships. This approach models mobility traces as contact graphs and identifies a set of seed users in both graphs, i.e., contacts graph and friendship in social network. In the second step, it propagates mapping from seed users to the remaining users in the graphs. This approach uses Distance Vector, Randomized Spanning Trees, and Recursive Subgraph Matching heuristics to measure the mapping strength and propagate the measured strength through the network.

Another work, from Ji et al. [85], improves the work of Srivasta et al. [168] in terms of accuracy and computational complexity. This work focuses on mapping anonymized users mobility traces to social media accounts. In addition to the users' local features, their approach incorporates users' global characteristics as well. Ji et al. define three similarity metrics: structural similarity, relative distance similarity, and inheritance similarity. These similarities are then combined in a unified similarity. Structural similarity considers features such as degree centrality, closeness centrality, and betweenness centrality while relative distance similarity captures the distance between users and seed users. Inheritance similarity considers the number of common neighbors that have been mapped as well as the degree similarity between the users in mobility traces and social media network graph. Next, Ji et al. [85] propose an adaptive de-anonymization framework that adaptively starts de-anonymizing from a core matching set that is consisted of a number of mapped users and k-hop mapping spanning set of them.

In another work, Reference [123], the location of Twitter users are inferred in different granularities (e.g., city, state, time zone, geographical region) based on their tweeting behavior (frequency of tweets per time unit) and the content of their tweets. This approach exploits external location knowledge (e.g., dictionary containing names of cities and states, and location-based services such as Foursquare) and finds explicit references of locations in tweets. Then all features are fed into a dynamically weighted method that is an ensemble of the statistical and heuristic classifiers.

Another work, from Wang et al. [177], links multiple users identities across multiple services/social media platforms (even with different types) according to the spatial-temporal locality of their activities, i.e., users mobility traces. This work also assumes that individuals can have multiple IDs/accounts. The motivation behind their algorithm is that IDs corresponding to the same person, are online at the same time in the same location and users' daily movement is predictable with repeated patterns. Wang et al. model users information as a contact graph where nodes are IDs (regardless of the service) and an edge represents connected IDs that have visited the same location. The weight of the edge demonstrates the number of co-location of two nodes. Then, a Bayesian matching algorithm is proposed to find the most probable matching candidates for a given target ID. A Bayesian inference method is then used to generate confidence scores for ranking candidates.

The work of Reference [91] compares different approaches in location inference attacks in social networks. There are also some other surveys discussing location inference techniques specifically in Twitter [7, 194], to which the reader can refer. Note that a large portion of research is dedicated to inference attacks on geolocated data, which is out of the scope of this survey [60, 112, 167]. A thorough survey is also available discussing geolocation data privacy, to which readers can refer if they are interested [112]. Note that the scope of this survey is a different from ours in which we cover the location privacy issues of users based on activities in social media.

In conclusion, location inference attack uses three types of information, (1) a user's network, (2) a user's contextual, and (3) a user's network and contextual information. A summary of the existing works is represented in Table 8 based on the type of leveraged information and used technique.

Table 8. Summary and Comparison of Different Location Inference Attacks

Technique \ Information	Network	Context	Network & Context
Probabilistic based	[20, 94, 126]	[44, 161, 177]	[108, 109]
Traditional supervised classification models (e.g., SVM, logistic regression, etc.)	[126, 160]	[72, 123]	-
Label propagation based	[47, 90, 168]	-	[85]

7 RECOMMENDATION SYSTEMS AND PRIVACY

Recommendation systems help individuals find information that matches with their interests by building user-interest profiles and recommending items to users based on those profiles. These profiles could be extracted from the users' interactions as they express their preferences and interests, e.g., clicks, likes/dislikes, ratings, purchases, and so on [25]. While user profiles help recommender systems to improve the quality of the services a user receives (a.k.a. utility), they also raise privacy concerns by reflecting the preferences of users [155]. Many works have studied the relationship between privacy and utility and have proposed solutions to handle the tradeoff. In general, these works focus on obfuscating users' interactions to hide their actual intentions and prevent accurate profiling [153, 157]. Following this strategy, no third parties or external entities need to be trusted by the users to preserve their privacy. Existing approaches use different techniques and mechanisms and could be categorized mainly into three categories: cryptographic-based techniques [6, 21, 40, 74, 172], differential privacy-based approaches [66, 76, 89, 120, 128, 130, 166, 197, 198], and perturbation-based techniques [75, 118, 146, 147, 148, 151, 153, 158, 183]

A group of works focus on providing cryptographic solutions to the problem of secure recommender systems. The approaches do not let the single trusted party have access to everyone's data [6, 21, 40, 74, 172]. Instead, users' ratings are stored as encrypted vectors and aggregates of the data are provided in the public domain. These approaches do not prevent privacy leaks through the output of recommendation systems (i.e., the recommendation themselves). These techniques are not the scope of this survey. Interested readers can refer to the mentioned papers for more details.

7.1 Differential Privacy-based Solutions

Works in this group utilize a differential privacy strategy to either anonymize user data before sending it to the recommendation system or perturb the recommendation outputs. McSherry et al. [128] modify leading algorithms for recommendation systems (i.e., SVD and k-nearest neighbor) for the first time so that drawing inferences about original ratings is difficult. They utilize differential privacy to construct private covariance matrices and make the collaborative filtering algorithms that use them private without having significant loss in accuracy.

In another work, Calandrino et al. [39] propose a new passive attack on recommender systems to infer a target user's transactions (i.e., item ratings). Their attack first monitors changes in the public outputs of a recommender system over a period of time. Public outputs may include related-items lists or an item-item covariance matrix. Then, it combines this information with a moderate amount of auxiliary information about the target user's transactions to further infer many of the target user's unknown transactions. Calandrino et al. further introduce an active inference attack on k-NN recommender systems. In this attack, k sybil users accounts are created and the k nearest neighbor of each sybil consists of $k-1$ other sybil users and the target user. The attack can then infer the target user's transactions history based on the items recommended to any of the sybils.

Results confirm the existence of privacy risks over the public outputs of recommender systems. The work of McSherry et al. [128] is not effective in protecting users against this attack as it does not consider updates to the covariance matrices and cannot provide a privacy guarantee in the dynamic settings. Machanavajjhala et al. [120] then quantifies the accuracy-privacy tradeoff. In particular, they prove lower bounds on the minimum loss in accuracy for recommendation systems that utilize differential privacy. Moreover, they adapt two differentially private algorithms, Laplace [53] and Exponential [129], to prevent disclosure of users' private attributes.

Previous works [120, 128] are vulnerable to k-nearest neighbor attack, as they fail to hide similar neighbors [39]. Zhu et al. [197] also propose a private neighborhood-based collaborative filtering that protects the information of both neighbors and user ratings. The proposed work assumes that the recommender system is trusted and introduces two operations: private neighbor selection and recommendation-aware sensitivity. The first operation seeks to protect neighbors identity by privately selecting k neighbors from a list of candidates and then adopting the exponential mechanism [129] to arrange a probability for each candidate. The second operation enhances the utility by reducing the magnitude of added noise. To do so, after selecting k neighbors, the similarity of neighbors is then perturbed by adding Laplace noise to mask the ratings given by a certain neighbor. Finally, the neighborhood collaborative filtering-based recommendation is performed on the private data. In another work, Jorgensen et al. [89] assume that all users' item-rating attributes are sensitive. However, different from Machanavajjhala et al. [120], they assume that users' social relations are non-sensitive. They propose a differentially private-based recommendation that incorporates social relations besides user-item ratings. To address the utility loss, this work first clusters users according to their social relations. Then, noisy averages of the user-item preferences are computed for each cluster using the differential privacy mechanism.

Shen et al. [166] assume that the recommender system is untrusted. They propose a user perturbation framework that anonymizes user data under a novel mechanism for differential privacy: relaxed admissible mechanism. The users' perturbed data is then used for recommendation. They provide mathematical bounds on the privacy and utility of the anonymized data. Hua et al. [76] also propose a differentially private matrix factorization-based recommender system. In particular, they solve this problem for two scenarios, trusted recommender and untrusted recommender. For the first scenario, user and item profile vectors are learned via regular and private version of matrix factorization, respectively. Private version of matrix factorization adds noises to item vectors to make them differentially private. In the second scenario, item profile vectors are first differentially privately learned with private matrix factorization problem. Then, a user's differentially private profile vector is derived from the private item profiles. A novel and strong form of differential privacy, namely, distance-based differential privacy, has been introduced by Guerroaui et al. [66]. Distance-based differential privacy ensures privacy for all the items rated by a user and the ones that are within a distance λ from it. The distance parameter λ controls the level of privacy and aids in tuning the recommendation privacy-utility tradeoff. The proposed protocol first finds a group of similar items for each given item. Then, it creates a manipulated user profile to preserve (ϵ, λ)-differential privacy by selecting an item and replacing it with another one.

Another differential privacy-based recommendation by Zhu et al. [198] proposed two approaches to solve the privacy problem in recommendation systems: item-based and user-based recommendation algorithms. In the item-based one, the exponential mechanism [129] is applied to the selection of the related items to guarantee differential privacy. Such resultant differentially private items list is further used to find recommendation for a given user. Similar private procedure happens in the user-based recommendation system. Another work differentiates sensitive and non-sensitive ratings to further improve the quality of recommendation systems in the long run [130]. Meng et al. [130] propose a personalized privacy-preserving recommender system.

Given sets of sensitive and non-sensitive ratings for each user, their approach utilizes differential privacy [52] to perturb users' ratings. Smaller and larger privacy budgets are considered for sensitive and non-sensitive ratings, respectively. This protects users' privacy while retaining recommendation effectiveness.

7.2 Perturbation-based Solutions

Perturbation-based techniques usually obfuscate users item ratings by adding random noise to the user data. Rebollo et al. [158] propose an approach that first measures the user's privacy risk as the KL divergence [48] between user's apparent profile and average population's distribution profile. The idea is that the more a user's profile diverges from the general population, the more information an attacker can learn about her. Then it seeks to find the obfuscation rate for generating forged user profiles so that the privacy risk is minimized. A closed-form solution is also provided for perturbing users interactions to optimize the privacy risk function.

Puglisi et al. [153] further extend Rebollo et al.'s work [158] to investigate the impact of this technique on content-based recommendation. This work measures a user's privacy risk similar to the approach proposed in Reference [158]. The utility is also measured by the prediction accuracy of the recommender system. It evaluates three different strategies, namely: optimized tag forgery [157], uniform tag forgery and TrackmeNot (TMN) [75]. The uniform tag forgery method assign forged tags according to a uniform distribution across all categories of the user profile. TMN constructed eleven categories from Open Directory Project (ODP) classification scheme[3] and selected the tags uniformly from this set. According to this work, users tend to mimic the profile of the population distribution when larger values of obfuscation rate is considered that results in less privacy risk but lower utility rate. Moreover, the authors have found that for a small forgery rate, it is possible to obtain an increase in privacy against a small degradation of utility.

Polat et al. [151] use a randomized perturbation technique [5] to obfuscate user generated data. Each user generates the disguised z-score for the item he has rated. The z-score for each user-item pair is based on the original item-rating, the user's average ratings and the total number of items she has rated. The proposed approach then passes the perturbed private data to the collaborative filtering-based recommender system. Another work, Reference [146], obfuscates user rating information and then passes disguised information to the collaborative filtering system for further recommendation. The proposed Nearest Neighbor Data Substitution (NeNDS) obfuscation method substitutes a user's data elements with one of her neighbors in the metric space [145]. However, one drawback of NeNDS is that the value of the perturbed data could be close enough to the original value that thus makes the data vulnerable. A hybrid version of NeNDS is then proposed that provides stronger privacy by geometrically transforming data before passing to NeNDS.

In contrast to Mcsherry et al. [128], Xin et al, assume that the recommender is not trusted and the onus is on the users to protect their privacy [183]. Their approach separates the computations that can be done by the users locally and privately and those that must be done by the recommender system. In particular, item features are learned by the system and user features are obtained locally by the users and further used for recommendation. Their approach also divides users into two groups, users who publicly share their information, and those who keep their preferences private. It then uses information of users in the first group to estimate items features. Xin et al. show theoretically and empirically that having the public information of a moderate number of users with a high number of ratings is enough to have an accurate estimation. Moreover, they propose a new privacy mechanism that privately releases second order information that is needed for estimating item features. This information is extracted from users who keep their preferences private. The

[3] http://www.dmoz.com.

main assumption behind this work is not realistic though, as in a real-world scenario it is not easy to collect ratings of a moderate number of people with a high number of ratings.

Luo et al. [118] propose a perturbation-based group recommendation method that assumes that similar users are grouped to each other and they are not willing to expose their preferences to anybody other than the group members. The items will be then recommended to the users within the same group. Particularly, in the first step, users are required to exchange their rating data among users in the same group given a secret key. This key varies for different users. The output of this step is a fake preference vector for each user. The value of the rating is then obfuscated in the second step by a chaos-based scrambling method. Similarly to Polat et al. [151], randomness is added to the output of the previous step to make sure no sensitive information remains in the published data. This information is then sent to the recommender system and it iteratively extracts information about aggregated ratings of the users. Extracted information is then used to estimate a group preference vector for collaborative filtering-based recommendation.

Parra-Arnau et al. [148], propose a privacy enhancing technology framework, PET, which perturbs users preferences information by combining two techniques, namely, the forgery and the suppression of ratings. In this scenario, users may avoid rating items they like and instead rate those that do not reflect their actual preferences. Therefore, the apparent profile of users will be different from their actual profile. Similarly to Reference [158], the privacy risk of each user is then measured as the KL divergence [48] between the user's apparent profile and the average population distribution. Utility is also controlled with the forgery and suppression rates. The tradeoff among privacy, forgery and suppression rate is then modeled as an optimization function to infer which ratings for each user should be forged and which ones should be suppressed to achieve the minimum privacy risk while keeping the utility of the data as high as possible. Similarly, Parra-Arnau et al. [147] propose a system that perturbs user rating profile according to her privacy preferences. The system has two components: (1) a profile-density model in which the user's profile will be more similar to the crowd's and (2) a classification model in which the user will not be identified as a member of a given group of users. The proposed model optimizes the tradeoff between privacy and utility and decides whether each service provider can have access to the user's profile, or not.

Recently, Biega et al. [33] proposed a framework that scrambles the users' rating history to preserve both their privacy and utility. The main assumption of this article is that recommender systems do not need the complete and accurate user profiles. Therefore, it splits users' profiles (i.e., pairs of user-item interactions) to *Mediator Accounts* (MA) in a way that coherent pieces of different users' profiles are kept intact in the MAs. The recommender will then deal with the MAs rather than real user profiles. This helps to preserve users' privacy by scrambling the user data across various proxy accounts while it keeps the user utility high as possible as well. Another work, from Guerraoui et al. [67], introduces metrics for measuring the utility and privacy effect of a user's behavior such as clicks, and likes/dislikes. Then, it shows that there is not always a tradeoff between utility and privacy. This article also proposed a click-advisor platform that warn users regarding the status of their click w.r.t. the privacy and utility. Here utility is defined as the difference between commonality of a user profile before and after a click. The privacy risk of a click is accordingly defined as the difference of the disclosure degree before and after the click.

Last, we summarize the reviewed state-of-the-art works in Table 9. Columns show the properties of the proposed models. These works protect user-item data (1) before sharing, (2) while processing the data, or (3) after recommending items to the user.

8 SUMMARY AND FUTURE RESEARCH DIRECTIONS

Explosive growth of the Web has not only drastically changed the way people conduct activities and acquire information but also has raised security [8, 14, 15] and privacy [26, 139] issues for

Table 9. Summary and Comparison of Different Anonymization Techniques in Recommendation Systems

		Privacy guarantee	Utility consideration	Input data anonymization	Middle data anonymization	Output data anonymization
Differential privacy	[128, 197]	✓		✓		
	[120]	✓	✓	✓		✓
	[66, 89, 166]	✓	✓	✓		
	[76]	✓		✓		✓
	[198]	✓			✓	✓
	[130]	✓	✓	✓	✓	✓
Perturbation	[67, 148, 153, 158]		✓	✓		
	[146, 151]			✓		
	[183]		✓	✓	✓	
	[118]			✓		✓
	[33]		✓	✓	✓	✓

them. Users are increasingly sharing their personal information on social media platforms. These platforms publish and share user-generated data with third-parties that risks exposing individuals' privacy. There are two general types of attacks, identity disclosure, and attribute disclosure. Sanitizing user-generated social media data is more challenging than structured data as it is heterogeneous, highly unstructured, noisy, and inherently different from relational and tabular data. In this survey, we review the recent developments in the field of privacy of social media data. We first review traditional privacy models for structural data. Then, we review, categorize, and compare existing methods in terms of privacy models, privacy leakage attacks, and anonymization algorithms. We also review privacy risks that exists in different aspects of social media such as users graph information, profile attributes, textual information and preferences. We categorize relevant works into five groups: (1) social graphs and privacy, (2) authors in social media and privacy, (3) profile attributes and privacy, (4) location and privacy, and (5) recommendation systems and privacy. For each category, we discuss existing attacks and solutions (if any was proposed) and classify them based on the type of data and the used technique. We outline the privacy attacks/solutions in Figure 1. Figure 2 also depicts the relevant privacy issues w.r.t. the type of social media data.

Detecting privacy issues and proposing techniques to protect users privacy in social media is a challenging issue. Most of the existing works focus on introducing new attacks and thus the gap between protection and detection becomes larger. Although a large body of work has emerged in recent years for investigating privacy issues for social media data, the development of tasks in each category is highly imbalanced. Some of them are well studied, whereas others need further investigation. We highlight these tasks in red in Figure 1 and Figure 2 based on privacy issues and user-generated data type, respectively. Below, we present some potential research directions:

- ***Protecting privacy of textual information***: Textual information is noisy, high-dimensional, and unstructured. It is rich in content and could reveal many sensitive information that user does not originally expose such as demographic information and location. This makes textual data a very important source of information for adversaries and could be exploited in many attacks. We thus need more research for anonymizing users' textual information to preserve privacy of users against various attacks such as author identification, and profile attribute disclosure.

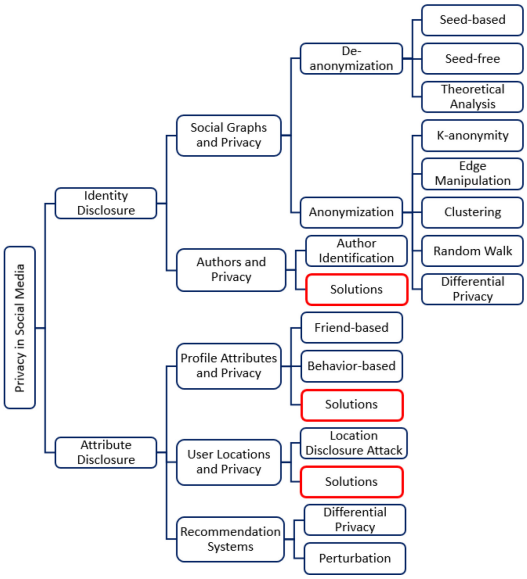

Fig. 1. An overview of privacy attacks and defenses in social media. Tasks highlighted in red have not been extensively studied.

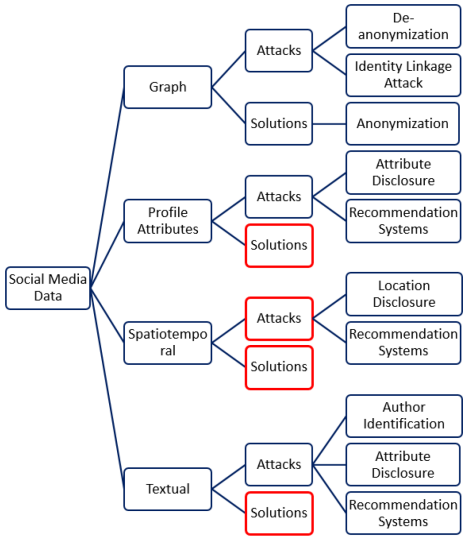

Fig. 2. An overview of privacy issues w.r.t. the type of social media data. Tasks highlighted in red have not been extensively studied.

- ***Protecting privacy of profile attribute information***: We reviewed many works that introduces privacy risks w.r.t. profile attributes. To the best of our knowledge, there is no work on introducing defense mechanisms against these attacks. One research direction could be in terms of a privacy-preserving tool for users that warns them against their activities and possibility of privacy leakage. Another direction is to propose a privacy protection technique that anonymize user's data before publishing to protect them against private attribute leakage.
- ***Privacy of spatiotemporal social media data***: Social media platforms support space-time indexed data and users have created a large volume of time-stamped, geo-located data. Such spatiotemporal data has an immense value for understanding users behavior better. In this survey, we review the state-of-the-art re-identification attacks that incorporate this data to breach privacy of users. This information may be used to infer users' location as well as their preferences and interests in case of recommendation systems. One future research direction could be investigating the role of temporal information in privacy of online users. More research should be done to build anonymization frameworks for protecting users temporal information.
- ***Privacy of heterogeneous social media data***: User-generated social media data is heterogeneous and consists of different aspects. Existing anonymization techniques assume that it is enough to anonymize each aspect of heterogeneous social media data independently. Beigi et al. [28] show that this assumption is not correct in practice due to the hidden relations between different aspects of the heterogeneous data. One potential research direction is to examine how different combinations of heterogeneous data (e.g., a combination of location and textual information) are vulnerable to the de-anonymization attack. Another potential direction is to improve anonymization techniques by considering hidden relations between different aspects of the data.

ACKNOWLEDGMENTS

The authors thank Alexander Nou for his help throughout the article.

REFERENCES

[1] Jemal H. Abawajy, Mohd Izuan Hafez Ninggal, and Tutut Herawan. 2016. Privacy preserving social network data publication. *IEEE Commun. Surv. Tutor.* 18, 3 (2016), 1974–1997.

[2] Ahmed Abbasi and Hsinchun Chen. 2008. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Trans. Inf. Syst.* 26, 2 (2008), 7.

[3] Sadia Afroz, Michael Brennan, and Rachel Greenstadt. 2012. Detecting hoaxes, frauds, and deception in writing style online. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy (SP'12)*. IEEE, 461–475.

[4] Gagan Aggarwal, Tomas Feder, Krishnaram Kenthapadi, Rajeev Motwani, Rina Panigrahy, Dilys Thomas, and An Zhu. 2005. Approximation algorithms for k-anonymity. In *Proceedings of the International Conference on Database Theory (ICDT'05)*.

[5] Rakesh Agrawal and Ramakrishnan Srikant. 2000. Privacy-preserving data mining. In *ACM SIGMOD Record*, Vol. 29.

[6] Esma Aimeur, Gilles Brassard, Jose M. Fernandez, Flavien Serge Mani Onana, and Zbigniew Rakowski. 2008. Experimental demonstration of a hybrid privacy-preserving recommender system. In *Availability, Reliability and Security*.

[7] Oluwaseun Ajao, Jun Hong, and Weiru Liu. 2015. A survey of location inference techniques on Twitter. *J. Inf. Sci.* 41, 6 (2015), 855–864.

[8] Muhammad Al-Qurishi, Mabrook Al-Rakhami, Atif Alamri, Majed Alrubaian, Sk Md Mizanur Rahman, and M Shamim Hossain. 2017. Sybil defense techniques in online social networks: A survey. *IEEE Access* 5 (2017), 1200–1219.

[9] Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. 2012. Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors. In *Sixth International AAAI Conference on Weblogs and Social Media (ICWSM'12)*.

[10] Mishari Almishari and Gene Tsudik. 2012. Exploring linkability of user reviews. In *Proceedings of the European Symposium on Research in Computer Security*. Springer, 307–324.

[11] Yasmeen Alufaisan, Yan Zhou, Murat Kantarcioglu, and Bhavani Thuraisingham. 2017. Hacking social network data mining. In *Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics (ISI'17)*. IEEE, 54–59.

[12] Hamidreza Alvari, Alireza Hajibagheri, Gita Sukthankar, and Kiran Lakkaraju. 2016. Identifying community structures in dynamic networks. *Soc. Netw. Anal. Min.* 6, 1 (2016), 77.

[13] Hamidreza Alvari, Kiran Lakkaraju, Gita Sukthankar, and Jon Whetzel. 2014. Predicting guild membership in massively multiplayer online games. In *Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 215–222.

[14] Hamidreza Alvari, Elham Shaabani, and Paulo Shakarian. 2018. Early identification of pathogenic social media accounts. In *Proceedings of the IEEE Intelligence and Security Informatics (ISI'18)*. IEEE.

[15] Hamidreza Alvari, Paulo Shakarian, and J. E. Kelly Snyder. 2017. Semi-supervised learning for detecting human trafficking. *Secur. Inf.* 6, 1 (2017), 1.

[16] Athanasios Andreou, Oana Goga, and Patrick Loiseau. 2017. Identity vs. attribute disclosure risks for users with multiple social profiles. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'17)*. ACM, 163–170.

[17] Michael Backes, Pascal Berrang, Oana Goga, Krishna P. Gummadi, and Praveen Manoharan. 2016. On profile linkability despite anonymity in social media systems. In *Proceedings of the ACM on Workshop on Privacy in the Electronic Society*.

[18] Michael Backes, Mathias Humbert, Jun Pang, and Yang Zhang. 2017. walk2friends: Inferring social links from mobility profiles. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*.

[19] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. 2007. Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th International Conference on the World Wide Web (WWW'07)*.

[20] Lars Backstrom, Eric Sun, and Cameron Marlow. 2010. Find me if you can: Improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th International Conference on the World Wide Web (WWW'10)*.

[21] Shahriar Badsha, Xun Yi, Ibrahim Khalil, and Elisa Bertino. 2017. Privacy preserving user-based recommender system. In *Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS'17)*. IEEE, 1074–1083.

[22] Ghazaleh Beigi. 2018. Social media and user privacy. *Arxiv Preprint Arxiv:1806.09786* (2018).
[23] Ghazaleh Beigi, Ruocheng Guo, Alexander Nou, Yanchao Zhang, and Huan Liu. 2019. Protecting user privacy: An approach for untraceable web browsing history and unambiguous user profiles. In *Proceedings of the 12th ACM International Conference on Web Search and Data Mining*. ACM, 213–221.
[24] Ghazaleh Beigi, Mahdi Jalili, Hamidreza Alvari, and Gita Sukthankar. 2014. Leveraging community detection for accurate trust prediction. In *Proceedings of the ASE International Conference on Social Computing*.
[25] Ghazaleh Beigi and Huan Liu. 2018. Similar but different: Exploiting users' congruity for recommendation systems. In *Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer.
[26] Ghazaleh Beigi and Huan Liu. 2019. Identifying novel privacy issues of online users on social media platforms by Ghazaleh Beigi and Huan Liu with Martin Vesely as coordinator. *ACM SIGWEB Newslett.* Article 4 (Winter, 2019), 7 pages. http://doi.acm.org/10.1145/3293874.3293878
[27] Ghazaleh Beigi, Suhas Ranganath, and Huan Liu. 2019. Signed link prediction with sparse data: The role of personality information. In *Companion Proceedings of the Web Conference 2019*. International World Wide Web Conferences Steering Committee.
[28] Ghazaleh Beigi, Kai Shu, Yanchao Zhang, and Huan Liu. 2018. Securing social media user data: An adversarial approach. In *Proceedings of the 29th Conference on Hypertext and Social Media*. ACM, 165–173.
[29] Ghazaleh Beigi, Jiliang Tang, and Huan Liu. 2016. Signed link analysis in social media networks. In *Proceedings of the 10th International Conference on Web and Social Media (ICWSM'16)*. AAAI Press.
[30] Ghazaleh Beigi, Jiliang Tang, Suhang Wang, and Huan Liu. 2016. Exploiting emotional information for trust/distrust prediction. In *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 81–89.
[31] Smriti Bhagat, Graham Cormode, Balachander Krishnamurthy, and Divesh Srivastava. 2009. Class-based graph anonymization for social network data. *Proc. VLDB Endow.* 2, 1 (2009), 766–777.
[32] Smriti Bhagat, Udi Weinsberg, Stratis Ioannidis, and Nina Taft. 2014. Recommending with an agenda: Active learning of private attributes using matrix factorization. In *Proceedings of the Recommender Systems Conference (RecSys'14)*. ACM.
[33] Asia J. Biega, Rishiraj Saha Roy, and Gerhard Weikum. 2017. Privacy through solidarity: A user-utility-preserving framework to counter profiling. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 665–674.
[34] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. In *Proceedings of Machine Learning Research (JMLR'03)*.
[35] Joseph Bonneau, Jonathan Anderson, and George Danezis. 2009. Prying data out of a social network. In *Proceedings of the International Conference on Advances in Social Network Analysis and Mining 2009 (ASONAM'09)*. IEEE, 249–254.
[36] Jasmine Bowers, Henry Williams, Gerry Dozier, and R Williams. 2015. Mitigation deanonymization attacks via language translation for anonymous social networks. *Proceedings of the International Conference on Machine Learning (ICML'15)* (2015).
[37] Joseph K. Bradley, Patrick Gage Kelley, and Aaron Roth. [n.d.]. Author identification from citations. ([n. d.]).
[38] Karl Bringmann, Tobias Friedrich, and Anton Krohmer. 2014. De-anonymization of heterogeneous random graphs in quasilinear time. In *Proceedings of the European Symposium on Algorithms*. Springer, 197–208.
[39] Joseph A. Calandrino, Ann Kilzer, Arvind Narayanan, Edward W. Felten, and Vitaly Shmatikov. 2011. " You might also like:" Privacy risks of collaborative filtering. In *Proceedings of the Security and Privacy (SP)*. IEEE.
[40] John Canny. 2002. Collaborative filtering with privacy via factor analysis. In *Proceedings of the SIGIR Conference on Research and Development in Information Retrieval*. ACM, 238–245.
[41] Abdelberi Chaabane, Gergely Acs, Mohamed Ali Kaafar, et al. 2012. You are what you like! information leakage through users' interests. In *Proceedings of the 19th Annual Network & Distributed System Security Symposium (NDSS)*.
[42] Carole E Chaski. 2005. Who is at the keyboard? Authorship attribution in digital evidence investigations. *International Journal of Digital Evidence* 4, 1 (2005), 1–13.
[43] James Cheng, Ada Wai-chee Fu, and Jia Liu. 2010. K-isomorphism: Privacy preserving network publication against structural attacks. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*.
[44] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You are where you tweet: A content-based approach to geo-locating twitter users. In *Proceedings of the Conference on Information and Knowledge Management (CIKM'10)*. ACM, 759–768.
[45] Carla-Fabiana Chiasserini, Michele Garetto, and Emilio Leonardi. 2016. Social network de-anonymization under scale-free user relations. *IEEE/ACM Trans. Netw.* 24, 6 (2016), 3756–3769.
[46] Carla-Fabiana Chiasserini, Michel Garetto, and Emili Leonardi. 2018. De-anonymizing clustered social networks by percolation graph matching. *ACM Trans. Knowl. Discov. Data* 12, 2 (2018), 21.

[47] Ryan Compton, David Jurgens, and David Allen. 2014. Geotagging one hundred million twitter accounts with total variation minimization. In *Proceedings of the 2014 IEEE International Conference on Big Data (Big Data'14)*. IEEE, 393–401.
[48] Thomas M. Cover and Joy A. Thomas. 2012. *Elements of Information Theory*. John Wiley & Sons.
[49] Ratan Dey, Cong Tang, Keith Ross, and Nitesh Saxena. 2012. Estimating age privacy leakage in online social networks. In *Proceedings of the 2012 Proceedings IEEE International Conference on Computer Communications (INFOCOM'12)*. IEEE, 2836–2840.
[50] Xenofontas Dimitropoulos, Dmitri Krioukov, Amin Vahdat, and George Riley. 2009. Graph annotations in modeling complex network topologies. *ACM Trans. Model. Comput. Simul.* 19, 4 (2009), 17.
[51] George T. Duncan and Diane Lambert. 1986. Disclosure-limited data dissemination. *J. Am. Stat. Assoc.* 81, 393 (1986), 10–18.
[52] Cynthia Dwork. 2008. Differential privacy: A survey of results. In *Proceedings of the International Conference on Theory and Applications of Models of Computation*. Springer, 1–19.
[53] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Theory of Cryptography Conference*. Springer, 265–284.
[54] Alexandre Evfimievski, Ramakrishnan Srikant, Rakesh Agrawal, and Johannes Gehrke. 2004. Privacy preserving mining of association rules. *Inf. Syst.* 29, 4 (2004), 343–364.
[55] Carla Fabiana, Michele Garetto, and Emilio Leonardi. 2015. De-anonymizing scale-free social networks by percolation graph matching. In *Proceedings of the 2015 IEEE International Conference on Computer Communications (INFOCOM'15)*. IEEE, 1571–1579.
[56] Hao Fu, Aston Zhang, and Xing Xie. 2014. De-anonymizing social graphs via node similarity. In *Proceedings of the Annual Conference on the World Wide Web (WWW'14)*.
[57] Hao Fu, Aston Zhang, and Xing Xie. 2015. Effective social graph deanonymization based on graph structure and descriptive information. *ACM Trans. Intell. Syst. Technol.* 6, 4 (2015), 49.
[58] Xinzhe Fu, Zhongzhao Hu, Zhiying Xu, Luoyi Fu, and Xinbing Wang. 2017. De-anonymization of networks with communities: When quantifications meet algorithms. In *Proceedings of the IEEE Global Communications Conference*.
[59] Benjamin C. M. Fung, K. Wang, R. Chen, and S. Yu Philip. 2010. Privacy-preserving data publishing: A survey on recent developments. *ACM Comput. Surv.* 42, 4 (2010), 1–53.
[60] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. 2010. Show me how you move and i will tell you who you are. In *Proceedings of the SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*.
[61] Daniel Gayo Avello. 2011. All liaisons are dangerous when all your friends are known to us. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*. ACM, 171–180.
[62] Oana Goga, Howard Lei, Sree Hari Krishnan Parthasarathi, Gerald Friedland, Robin Sommer, and Renata Teixeira. 2013. Exploiting innocuous activity for correlating users across sites. In *Proceedings of the Annual Conference on the World Wide Web (WWW'13)*.
[63] Neil Zhenqiang Gong and Bin Liu. 2016. You are who you know and how you behave: Attribute inference attacks via users' social friends and behaviors. In *Proceedings of the USENIX Security Symposium*. 979–995.
[64] Neil Zhenqiang Gong and Bin Liu. 2018. Attribute inference attacks in online social networks. *ACM Trans. Priv. Secur.* 21, 1 (2018), 3.
[65] Neil Zhenqiang Gong, Ameet Talwalkar, Lester Mackey, Ling Huang, Eui Chul Richard Shin, Emil Stefanov, Elaine Runting Shi, and Dawn Song. 2014. Joint link prediction and attribute inference using a social-attribute network. *ACM Trans. Intell. Syst. Technol.* 5, 2 (2014), 27.
[66] Rachid Guerraoui, Anne-Marie Kermarrec, Rhicheek Patra, and Mahsa Taziki. 2015. D 2 p: Distance-based differential privacy in recommenders. *Proc. VLDB Endow.* 8, 8 (2015), 862–873.
[67] Rachid Guerraoui, Anne-Marie Kermarrec, and Mahsa Taziki. 2017. The utility and privacy effects of a click. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.
[68] Payas Gupta, Swapna Gottipati, Jing Jiang, and Debin Gao. 2013. Your love is public now: Questioning the use of personal information in authentication. In *Proceedings of the ACM Special Interest Group on Security, Audit and Control Conference (SIGSAC'13)*. ACM.
[69] Alireza Hajibagheri, Gita Sukthankar, Kiran Lakkaraju, Hamidreza Alvari, Rolf T. Wigand, and Nitin Agarwal. 2018. Using massively multiplayer online game data to analyze the dynamics of social interactions. *Social Interactions in Virtual Worlds: An Interdisciplinary Perspective* (2018).
[70] Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis. 2008. Resisting structural re-identification in anonymized social networks. *Proc. VLDB Endow.* 1, 1 (2008), 102–114.
[71] Jianming He, Wesley W. Chu, and Zhenyu Victor Liu. 2006. Inferring privacy information from social networks. In *Proceedings of the International Conference on Intelligence and Security Informatics*. Springer, 154–165.

[72] Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H. Chi. 2011. Tweets from justin bieber's heart: The dynamics of the location field in user profiles. In *Proceedings of the Conference of the Special Interest Group on Computer-Human Interaction (SIGCHI'11)*. ACM, 237–246.

[73] Shawndra Hill and Foster Provost. 2003. The myth of the double-blind review?: Author identification using only citations. *ACM SIGKDD Explor. Newslett.* 5, 2 (2003), 179–184.

[74] T. Ryan Hoens, Marina Blanton, and Nitesh V. Chawla. 2010. A private and reliable recommendation system for social networks. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing (SocialCom'10)*. IEEE, 816–825.

[75] D. C. Howe and H. Nissenbaum. 2009. TrackMeNot: Resisting surveillance in web search. In *Lessons from the Identity Trail: Privacy, Anonymity and Identity in a Networked Society*. (Oxford University Press, New York, 2009), 417–436.

[76] Jingyu Hua, Chang Xia, and Sheng Zhong. 2015. Differentially private matrix factorization. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'15)*.

[77] Mathias Humbert, Théophile Studer, Matthias Grossglauser, and Jean-Pierre Hubaux. 2013. Nowhere to hide: Navigating around privacy in online social networks. In *Proceedings of the European Symposium on Research in Computer Security*.

[78] Piotr Indyk and Rajeev Motwani. 1998. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*. ACM, 604–613.

[79] P. James. 1992. Knowledge graphs. *Linguistic Instruments in Knowledge Engineering* (1992), 97–117.

[80] Shouling Ji, Weiqing Li, Neil Zhenqiang Gong, Prateek Mittal, and Raheem A. Beyah. 2015. On your social network de-anonymizablity: Quantification and large scale evaluation with seed knowledge. In *Proceedings of the Network and Distributed System Security Symposium (NDSS'15)*.

[81] Shouling Ji, Weiqing Li, Neil Zhenqiang Gong, Prateek Mittal, and Raheem A. Beyah. 2016. Seed based deanonymizability quantification of social networks. *IEEE Trans. Inf. Forens. Secur.* 11, 7, 1398–1411.

[82] Shouling Ji, Weiqing Li, Prateek Mittal, and Raheem Beyah. 2015. SecGraph: A uniform and open-source evaluation system for graph data anonymization and de-anonymization. In *Proceedings of the USENIX Security Symposium*. 303–318.

[83] Shouling Ji, Weiqing Li, Mudhakar Srivatsa, and Raheem Beyah. 2014. Structural data de-anonymization: Quantification, practice, and implications. In *Proceedings of the 2014 ACM Special Interest Group on Security, Audit and Control Conference (SIGSAC'14)*. ACM, 1040–1053.

[84] Shouling Ji, Weiqing Li, Mudhakar Srivatsa, and Raheem Beyah. 2016. Structural data de-anonymization: Theory and practice. *IEEE/ACM Trans. Netw.* 24, 6 (2016), 3523–3536.

[85] Shouling Ji, Weiqing Li, Mudhakar Srivatsa, Jing Selena He, and Raheem Beyah. 2016. General graph data de-anonymization: From mobility traces to social networks. *ACM Trans. Intell. Syst. Technol.* 18, 4 (2016).

[86] Shouling Ji, Prateek Mittal, and Raheem Beyah. 2016. Graph data anonymization, de-anonymization attacks, and de-anonymizability quantification: A survey. *IEEE Commun. Surv. Tutor.* 19, 2 (2016), 1305–1326.

[87] Shouling Ji, Ting Wang, Jianhai Chen, Weiqing Li, Prateek Mittal, and Raheem Beyah. 2017. De-SAG: On the de-anonymization of structure-attribute graph data. *IEEE Trans. Depend. Sec. Comput.* 16, 4 (2017), 594–607.

[88] Jinyuan Jia, Binghui Wang, Le Zhang, and Neil Zhenqiang Gong. 2017. AttriInfer: Inferring user attributes in online social networks using markov random fields. In *Proceedings of the Annual Conference on the world wide web (WWW'17)*. 1561–1569.

[89] Zach Jorgensen and Ting Yu. 2014. A privacy-preserving framework for personalized, social recommendations. In *Proceedings of the Extended Database Technology Conference (EDBT'14)*. 582.

[90] David Jurgens. 2013. That's what friends are for: Inferring location in online social media platforms based on social relationships. In *Seventh International AAAI Conference on Weblogs and Social Media*.

[91] David Jurgens, Tyler Finethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015. Geolocation prediction in twitter using social networks: A critical analysis and review of current practice. In *Ninth International AAAI Conference on Web and Social Media*.

[92] Daniel Kifer and Ashwin Machanavajjhala. 2011. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*. ACM, 193–204.

[93] Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*.

[94] Longbo Kong, Zhi Liu, and Yan Huang. 2014. Spot: Locating social media users based on social network context. *Proc. VLDB Endow.* 7, 13 (2014), 1681–1684.

[95] Moshe Koppel, Jonathan Schler, and Shlomo Argamon. 2009. Computational methods in authorship attribution. *J. Assoc. Inf. Sci. Technol.* 60, 1 (2009), 9–26.

[96] Moshe Koppel, Jonathan Schler, and Shlomo Argamon. 2011. Authorship attribution in the wild. *Lang. Resourc. Eval.* 45, 1 (2011), 83–94.

[97] Moshe Koppel, Jonathan Schler, Shlomo Argamon, and Eran Messeri. 2006. Authorship attribution with thousands of candidate authors. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 659–660.

[98] Aleksandra Korolova, Rajeev Motwani, Shubha U Nabar, and Ying Xu. 2008. Link privacy in social networks. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. ACM, 289–298.

[99] Nitish Korula and Silvio Lattanzi. 2014. An efficient reconciliation algorithm for social networks. *Proc. VLDB Endow.* 7, 5 (2014), 377–388.

[100] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15 (2013), 5802–5805.

[101] Harold W. Kuhn. 2010. The hungarian method for the assignment problem. In *50 Years of Integer Programming 1958-2008*. Springer, 29–47.

[102] Sebastian Labitzke, Florian Werling, Jens Mittag, and Hannes Hartenstein. 2013. Do online social network friends still threaten my privacy? In *Proceedings of the ACM Conference on Data and Application Security and Privacy*.

[103] Diane Lambert. 1993. Measures of disclosure risk and harm. *J. Off. Stat.* 9, 2 (1993), 313.

[104] Wei-Han Lee, Changchang Liu, Shouling Ji, Prateek Mittal, and Ruby B. Lee. 2017. How to quantify graph de-anonymization risks. In *International Conference on Information Systems Security and Privacy*. Springer, 84–104.

[105] Wei-Han Lee, Changchang Liu, Shouling Ji, Prateek Mittal, and Ruby B. Lee. 2017. Blind de-anonymization attacks using social networks. In *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society*. ACM, 1–4.

[106] Kevin Lewis, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. 2008. Tastes, ties, and time: A new social network dataset using facebook. com. *Soc. Netw.* 30, 4 (2008), 330–342.

[107] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. 2007. t-closeness: Privacy beyond k-anonymity and l-diversity. In *Proceedings of the IEEE 23rd International Conference on Data Engineering 2007 (ICDE'07)*. IEEE, 106–115.

[108] Rui Li, Shengjie Wang, and Kevin Chen-Chuan Chang. 2012. Multiple location profiling for users and relationships from social network and content. *Proc. VLDB Endow.* 5, 11 (2012), 1603–1614.

[109] Rui Li, Shengjie Wang, Hongbo Deng, Rui Wang, and Kevin Chen-Chuan Chang. 2012. Towards social user profiling: Unified and discriminative influence model for inferring home locations. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD'12)*.

[110] Xiaoxue Li, Yanan Cao, Yanmin Shang, Yanbing Liu, Jianlong Tan, and Li Guo. 2017. Inferring user profiles in online social networks based on convolutional neural network. In *Proceedings of the International Conference on Knowledge Science, Engineering and Management*. Springer.

[111] Jack Lindamood, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. 2009. Inferring private information using social network data. In *Proceedings of the Annual Conference of the World Wide Web (WWW'09)*. ACM, 1145–1146.

[112] Bo Liu, Wanlei Zhou, Tianqing Zhu, Longxiang Gao, and Yong Xiang. 2018. Location privacy and its applications: A systematic study. *IEEE Access* 6 (2018), 17606–17624.

[113] Changchang Liu, Supriyo Chakraborty, and Prateek Mittal. 2016. Dependence makes you vulnberable: Differential privacy under dependent tuples. In *Proceedings of the Network and Distributed System Security Symposium (NDSS'16)*, Vol. 16. 21–24.

[114] Changchang Liu and Prateek Mittal. 2016. LinkMirage: Enabling privacy-preserving analytics on social relationships. In *Proceedings of the Network and Distributed System Security Symposium (NDSS'16)*.

[115] Kun Liu and Evimaria Terzi. 2008. Towards identity anonymization on graphs. In *Proceedings of the ACM Special Interest Group on Management of Data Conference (SIGMOD'08)*.

[116] Yushan Liu, Shouling Ji, and Prateek Mittal. 2016. SmartWalk: Enhancing social network security via adaptive random walks. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 492–503.

[117] Dixin Luo, Hongteng Xu, Hongyuan Zha, Jun Du, Rong Xie, Xiaokang Yang, and Wenjun Zhang. 2014. You are what you watch and when you watch: Inferring household structures from iptv viewing data. *IEEE Trans. Broadcast.* 60, 1 (2014), 61–72.

[118] Zhifeng Luo and Zhanli Chen. 2014. A privacy preserving group recommender based on cooperative perturbation. In *Proceedings of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. IEEE.

[119] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, and Muthuramakrishnan Venkitasubramaniam. 2006. l-diversity: Privacy beyond k-anonymity. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE'06)*. IEEE, 24–24.

[120] Ashwin Machanavajjhala, Aleksandra Korolova, and Atish Das Sarma. 2011. Personalized social recommendations: Accurate or private. *Proc. VLDB Endow.* 4, 7 (2011), 440–450.

[121] Nathan Mack, Jasmine Bowers, Henry Williams, Gerry Dozier, and Joseph Shelton. 2015. The best way to a strong defense is a strong offense: Mitigating deanonymization attacks via iterative language translation. *Int. J. Mach. Learn. Comput.* 5, 5 (2015), 409.

[122] Priya Mahadevan, Dmitri Krioukov, Kevin Fall, and Amin Vahdat. 2006. Systematic topology analysis and generation using degree correlations. In *Proceedings of the ACM SIGCOMM Computer Communication Review*, Vol. 36. ACM, 135–146.

[123] Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. 2014. Home location identification of twitter users. *ACM Trans. Intell. Syst. Technol.* 5, 3 (2014), 47.

[124] Huina Mao, Xin Shuai, and Apu Kapadia. 2011. Loose tweets: An analysis of privacy leaks on twitter. In *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society*. ACM, 1–12.

[125] Jeffrey McGee, James Caverlee, and Zhiyuan Cheng. 2013. Location prediction in social media based on tie strength. In *Proceedings of the Conference on Information and Knowledge Management (CIKM'13)*. ACM.

[126] Jeffrey McGee, James A. Caverlee, and Zhiyuan Cheng. 2011. A geographic study of tie strength in social media. In *Proceedings of the Conference on Information and Knowledge Management (CIKM'11)*. ACM, 2333–2336.

[127] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* 27, 1 (2001), 415–444.

[128] Frank McSherry and Ilya Mironov. 2009. Differentially private recommender systems: Building privacy into the netflix prize contenders. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD'09)*. ACM.

[129] Frank McSherry and Kunal Talwar. 2007. Mechanism design via differential privacy. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science 2007 (FOCS'07)*. IEEE, 94–103.

[130] Xuying Meng, Suhang Wang, Kai Shu, Jundong Li, Bo Chen, Huan Liu, and Yujun Zhang. 2018. Personalized privacy-preserving social recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'18)*.

[131] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*. 3111–3119.

[132] Tehila Minkus, Yuan Ding, Ratan Dey, and Keith W. Ross. 2015. The city privacy attack: Combining social media and public records for detailed profiles of adults and children. In *Proceedings of the ACM Conference on Online Social Networks*.

[133] Alan Mislove, Bimal Viswanath, Krishna P. Gummadi, and Peter Druschel. 2010. You are who you know: Inferring user profiles in online social networks. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM'10)*. ACM, 251–260.

[134] Prateek Mittal, Charalampos Papamanthou, and Dawn Song. 2013. Preserving link privacy in social network based systems. *NDSS*.

[135] Frederick Mosteller and David Wallace. 1964. *Inference and Disputed Authorship: The Federalist*. Addison-Wesley, Reading, Mass.

[136] Mihir Nanavati, Nathan Taylor, William Aiello, and Andrew Warfield. 2011. Herbert west-deanonymizer. In *Proceedings of the 6th USENIX Conference on Hot Topics in Security (HotSec'11)*. USENIX Association, San Francisco, CA, 6–6.

[137] Arvind Narayanan, Hristo Paskov, Neil Zhenqiang Gong, John Bethencourt, Emil Stefanov, Eui Chul Richard Shin, and Dawn Song. 2012. On the feasibility of internet-scale author identification. In *Proceedings of the Conference on Security and Privacy (SP'12)*. IEEE.

[138] Arvind Narayanan, Elaine Shi, and Benjamin IP Rubinstein. 2011. Link prediction by de-anonymization: How we won the kaggle social network challenge. In *Proceedings of the International Joint Conference on Neural Networks*. IEEE.

[139] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust de-anonymization of large sparse datasets. In *Proceedings of the Conference on Security and Privacy*. IEEE.

[140] Arvind Narayanan and Vitaly Shmatikov. 2009. De-anonymizing social networks. In *Proceedings of the Conference on Security and Privacy*. IEEE.

[141] M. E. J. Newman. 2003. The structure and function of complex networks. In *SIAM Review*, Vol. 45. 167–256.

[142] Shirin Nilizadeh, Apu Kapadia, and Yong-Yeol Ahn. 2014. Community-enhanced de-anonymization of online social networks. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 537–548.

[143] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2007. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*. ACM, 75–84.

[144] Jahna Otterbacher. 2010. Inferring gender of movie reviewers: Exploiting writing style, content and metadata. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. ACM, 369–378.

[145] Rupa Parameswaran and D. Blough. 2005. A robust data obfuscation approach for privacy preservation of clustered data. In *Proceedings of the Workshop on Privacy and Security Aspects of Data Mining*. 18–25.

[146] Rupa Parameswaran and Douglas M. Blough. 2007. Privacy preserving collaborative filtering using data obfuscation. In *Proceedings of the IEEE International Conference on Granular Computing*.

[147] Javier Parra-Arnau. 2017. Pay-per-tracking: A collaborative masking model for web browsing. *Information Sciences* 385–386 (2017), 96–124.

[148] Javier Parra-Arnau, David Rebollo-Monedero, and Jordi Forné. 2014. Optimal forgery and suppression of ratings for privacy enhancement in recommendation systems. *Entropy* 16, 3 (2014), 1586–1631.

[149] Pedram Pedarsani and Matthias Grossglauser. 2011. On the privacy of anonymized networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1235–1243.

[150] Wei Peng, Feng Li, Xukai Zou, and Jie Wu. 2014. A two-stage deanonymization attack against anonymized social networks. *IEEE Trans. Comput.* 63, 2 (2014), 290–303.

[151] Huseyin Polat and Wenliang Du. 2003. Privacy-preserving collaborative filtering using randomized perturbation techniques. In *Proceedings of the 3rd IEEE International Conference on Data Mining 2003 (ICDM'03)*. IEEE, 625–628.

[152] Davide Proserpio, Sharon Goldberg, and Frank McSherry. 2014. Calibrating data to sensitivity in private data analysis: A platform for differentially-private analysis of weighted datasets. *Proc. VLDB Endow.* 7, 8 (2014).

[153] Silvia Puglisi, Javier Parra-Arnau, Jordi Forné, and David Rebollo-Monedero. 2015. On content-based recommendation and user privacy in social-tagging systems. *Comput. Stand. Interfaces* 41 (2015), 17–27.

[154] Jianwei Qian, Xiang-Yang Li, Chunhong Zhang, and Linlin Chen. 2016. De-anonymizing social networks and inferring private attributes using knowledge graphs. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM'16)*.

[155] Naren Ramakrishnan, Benjamin J. Keller, Batul J. Mirza, Ananth Y. Grama, and George Karypis. 2001. Privacy risks in recommender systems. *IEEE Internet Comput.* 5, 6 (2001), 54.

[156] Josyula R. Rao, Pankaj Rohatgi, et al. 2000. Can pseudonymity really guarantee privacy? In *Proceedings of the USENIX Conference on Security*.

[157] David Rebollo-Monedero and Jordi Forné. 2010. Optimized query forgery for private information retrieval. *IEEE Trans. Inf. Theory* 56, 9 (2010), 4631–4642.

[158] David Rebollo-Monedero, Javier Parra-Arnau, and Jordi Forné. 2011. An information-theoretic privacy criterion for query forgery in information retrieval. In *Proceedings of the International Conference on Security Technology*. Springer, 146–154.

[159] Shariq J Rizvi and Jayant R Haritsa. 2002. Maintaining data privacy in association rule mining. In *Proceedings of the 28th International Conference on Very Large Databases (VLDB'02)*. Elsevier, 682–693.

[160] Dominic Rout, Kalina Bontcheva, Daniel Preoţiuc-Pietro, and Trevor Cohn. 2013. Where's@ wally?: A classification approach to geolocating users based on their social ties. In *Proceedings of the Annual Conference on Hypertext and Social Media*. ACM.

[161] KyoungMin Ryoo and Sue Moon. 2014. Inferring twitter user locations with 10 km accuracy. In *Proceedings of the Annual Conference on the World Wide Web (WWW'14)*.

[162] Alessandra Sala, Xiaohan Zhao, Christo Wilson, Haitao Zheng, and Ben Y Zhao. 2011. Sharing graphs using differentially private graph models. In *Proceedings of the ACM SIGCOMM on Internet Measurement Conference*.

[163] Kumar Sharad. 2016. Change of guard: The next generation of social graph de-anonymization attacks. In *Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security*. ACM, 105–116.

[164] Kumar Sharad and George Danezis. 2014. An automated social graph de-anonymization technique. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*. ACM, 47–58.

[165] Sanur Sharma, Preeti Gupta, and Vishal Bhatnagar. 2012. Anonymisation in social network: A literature survey and classification. *Int. J. Soc. Netw. Min.* 1, 1 (2012), 51–66.

[166] Yilin Shen and Hongxia Jin. 2014. Privacy-preserving personalized recommendation: An instance-based approach via differential privacy. In *Proceedings of the 2014 IEEE International Conference on Data Mining (ICDM'14)*. IEEE, 540–549.

[167] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. 2011. Quantifying location privacy. In *Proceedings of the 2011 IEEE Symposium on Security and Privacy (SP'11)*. IEEE, 247–262.

[168] Mudhakar Srivatsa and Mike Hicks. 2012. Deanonymizing mobility traces: Using social network as a side-channel. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*. ACM, 628–637.

[169] Efstathios Stamatatos. 2009. A survey of modern authorship attribution methods. *J. Assoc. Inf. Sci. Technol.* 60, 3 (2009), 538–556.

[170] Zak Stone, Todd Zickler, and Trevor Darrell. 2008. Autotagging facebook: Social network context improves photo annotation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*.

[171] Latanya Sweeney. 2002. k-anonymity: A model for protecting privacy. *Int. J. Uncert. Fuzz. Knowl.-Based Syst.* 10, 05 (2002), 557–570.

[172] Qiang Tang and Jun Wang. 2018. Privacy-preserving friendship-based recommender systems. *IEEE Trans. Depend. Sec. Comput.* 15, 5 (2018), 784–796.

[173] Kurt Thomas, Chris Grier, and David M Nicol. 2010. unfriendly: Multi-party privacy risks in social networks. In *Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 236–252.

[174] Brian Thompson and Danfeng Yao. 2009. The union-split algorithm and cluster-based anonymization of social networks. In *Proceedings of the Symposium on Information, Computer, and Communications Security*.

[175] Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan. 2006. Fast random walk with restart and its applications. In *Proceedings of the Sixth International Conference on Data Mining (ICDM'06)*. IEEE Computer Society, 613–622.

[176] Vassilios S. Verykios, Elisa Bertino, Igor Nai Fovino, Loredana Parasiliti Provenza, Yucel Saygin, and Yannis Theodoridis. 2004. State-of-the-art in privacy preserving data mining. *ACM SIGMOD Rec.* 33, 1 (2004), 50–57.

[177] Huandong Wang, Yong Li, Gang Wang, and Depeng Jin. 2018. You are how you move: Linking multiple user identities from massive mobility traces. In *Proceedings of the SIAM International Conference on Data Mining (SDM'18)*. Society for Industrial and Applied Mathematics.

[178] Pengfei Wang, Jiafeng Guo, Yanyan Lan, Jun Xu, and Xueqi Cheng. 2016. Your cart tells you: Inferring demographic attributes from purchase data. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM'16)*. ACM.

[179] Yue Wang and Xintao Wu. 2013. Preserving differential privacy in degree-correlation based graph generation. *Trans. Data Priv.* 6, 2 (2013), 127.

[180] Udi Weinsberg, Smriti Bhagat, Stratis Ioannidis, and Nina Taft. 2012. BlurMe: Inferring and obfuscating user gender based on ratings. In *Proceedings of the 6th ACM Conference on Recommender Systems*. ACM, 195–202.

[181] Xinyu Wu, Zhongzhao Hu, Xinzhe Fu, Luoyi Fu, Xinbing Wang, and Songwu Lu. 2018. Social network deanonymization with overlapping communities: Analysis, algorithm and experiments. In *Proceeding of the International Conference on Computer Communications (INFOCOM'18)*.

[182] Qian Xiao, Rui Chen, and Kian-Lee Tan. 2014. Differentially private network data release via structural inference. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 911–920.

[183] Yu Xin and Tommi Jaakkola. 2014. Controlling privacy in recommender systems. In *Proceedings of the Conference on Neural Information Processing Systems (NIPS'14)*.

[184] Jaewon Yang and Jure Leskovec. 2013. Overlapping community detection at scale: A nonnegative matrix factorization approach. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*. ACM, 587–596.

[185] Lyudmila Yartseva and Matthias Grossglauser. 2013. On the performance of percolation graph matching. In *Proceedings of the 1st ACM Conference on Online Social Networks*. ACM, 119–130.

[186] Zhijun Yin, Manish Gupta, Tim Weninger, and Jiawei Han. 2010. Linkrec: A unified framework for link recommendation with user attributes and graph structure. In *Proceedings of the Annual Conference of the World Wide Web (WWW'10)*. ACM, 1211–1212.

[187] Zhijun Yin, Manish Gupta, Tim Weninger, and Jiawei Han. 2010. A unified framework for link recommendation using random walks. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM'10)*. IEEE, 152–159.

[188] Xiaowei Ying and Xintao Wu. 2009. Graph generation with prescribed feature constraints. In *Proceedings of the SIAM International Conference on Data Mining (SDM'09)*.

[189] Mingxuan Yuan, Lei Chen, and Philip S. Yu. 2010. Personalized privacy protection in social networks. *Proc. VLDB Endow.* 4, 2 (2010), 141–150.

[190] Aston Zhang, Xing Xie, Carl A. Gunter, Jiawei Han, and XiaoFeng Wang. 2014. Privacy risk in anonymized heterogeneous information networks. In *Proceedings of the Extended Database Technology Conference (EDBT'14)*.

[191] Jinxue Zhang, Jingchao Sun, Rui Zhang, and Yanchao Zhang. 2018. Privacy-preserving social media data outsourcing. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM'18)*.

[192] Elena Zheleva and Lise Getoor. 2009. To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. In *Proceedings of the 18th International Conference on World Wide Web*. ACM, 531–540.

[193] Elena Zheleva, Evimaria Terzi, and Lise Getoor. 2012. Privacy in social networks. *Synth. Lect. Data Min. Knowl. Discov.* 3, 1 (2012), 1–85.

[194] X. Zheng, J. Han, and A. Sun. 2018. A survey of location prediction on twitter. *IEEE Trans. Knowl. Data Eng.* 30, 9 (2018), 1652–1671.

[195] Yuan Zhong, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang, and Xing Xie. 2015. You are where you go: Inferring demographic attributes from location check-ins. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM'15)*. ACM, 295–304.

[196] Bin Zhou and Jian Pei. 2008. Preserving privacy in social networks against neighborhood attacks. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE'08)*.

[197] Tianqing Zhu, Gang Li, Yongli Ren, Wanlei Zhou, and Ping Xiong. 2013. Differential privacy for neighborhood-based collaborative filtering. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13)*. ACM, 752–759.

[198] Xue Zhu and Yuqing Sun. 2016. Differential privacy for collaborative filtering recommender algorithm. In *Proceedings of the 2016 ACM on International Workshop on Security And Privacy Analytics*. ACM, 9–16.

[199] Lei Zou, Lei Chen, and M. Tamer Özsu. 2009. K-automorphism: A general framework for privacy preserving network publication. *Proc. VLDB Endow.* 2, 1 (2009), 946–957.

Received July 2018; revised January 2019; accepted April 2019