Journal of Privacy and Confidentiality
Vol. 14 (1) 2023

Submitted May 2022
Published Feb 2024

DIFFERENTIAL PRIVACY GUARANTEES FOR ANALYTICS AND MACHINE LEARNING ON GRAPHS: A SURVEY OF RESULTS

TAMARA T. MUELLER, DMITRII USYNIN, JOHANNES C. PAETZOLD, RICKMER BRAREN, DANIEL RUECKERT, AND GEORGIOS KAISSIS

Institute for Artificial Intelligence in Medicine and Healthcare and Institute of Radiology, Technical University of Munich
e-mail address: tamara.mueller@tum.de

Institute for Artificial Intelligence in Medicine and Healthcare and Institute of Radiology, Technical University of Munich

Department of Informatics, Technical University of Munich; Institute for Tissue Engineering and Regenerative Medicine, Helmholtz Zentrum München

Institute of Radiology, Technical University of Munich

Institute for Artificial Intelligence in Medicine and Healthcare and Institute of Radiology, Technical University of Munich; Department of Computing, Imperial College London

Institute for Artificial Intelligence in Medicine and Healthcare and Institute of Radiology, Technical University of Munich; Department of Computing, Imperial College London, Helmholtz Zentrum München

Abstract. We study differential privacy (DP) in the context of graph-structured data and discuss its formulations and applications to the publication of graphs and their associated statistics, graph generation methods, and machine learning on graph-based data, including graph neural networks (GNNs). Interpreting DP guarantees in the context of graph-structured data can be challenging, as individual data points are interconnected (often non-linearly or sparsely). This differentiates graph databases from tabular databases, which are usually used in DP, and complicates related concepts like the derivation of per-sample gradients in GNNs. The problem is exacerbated by an absence of a single, well-established formulation of DP in graph settings. A lack of prior systematisation work motivated us to study graph-based learning from a privacy perspective. In this work, we systematise different formulations of DP on graphs, and discuss challenges and promising applications, including the GNN domain. We compare and separate works into methods that privately estimate graph data (either by statistical analysis or using GNNs), and methods that aim at generating new graph data. We conclude our work with a discussion of open questions and potential directions for further research in this area.

Key words and phrases: differential privacy, graph-structured data, graph analytics, graph neural networks.

JPrivacy Confidentiality www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.820

© T.T. Mueller, D. Usynin, J.C. Paetzold, R. Braren, D. Rueckert, and G. Kaissis
Creative Commons (CC BY-NC-ND 4.0)

1. Introduction

Many real-world datasets like social networks, molecules, population data or electronic health records do not naturally befit a row-and-column (tabular) representation as they hold complex internal connections and relationships. Such data can often be efficiently represented using graphs as data structures. The additional intrinsic structural information maintained by this representation holds great potential for data analytics and learning tasks on such graph-structured data. A graph's interconnected nature can be leveraged by appropriate algorithms and graph-based learning models and can be deployed in contexts such as market value prediction [77], fake news detection [7] and drug development [36]. Within the last two decades, "traditional" algorithms such as triangle counting, node degree estimation etc. have been complemented or superseded by advanced machine learning applications on graph-structured data, made possible by the introduction of graph neural networks (GNNs) [101]. Such models have since then been successfully applied to various learning scenarios [47, 85, 98]. These works demonstrate that a graph's connectivity confers valuable additional information, and allows analysts to leverage the interaction between individual data points, which can significantly improve the accuracy of learning tasks compared to reducing graph-structured data to a tabular form [121]. However, the information contained in graph-structured data is often highly sensitive in nature in the sense that either the data in the graph's *nodes*, the *connections between nodes* or *both* represent sensitive information requiring protection.

Moreover, the rich inter-node relationships render graph-structured data more vulnerable to attacks that attempt do disclose the private data of individuals contained within the graph without their consent [68, 132]. Such attacks can take the form of membership inference (MIA) [104], where the adversary attempts to verify if a record that they possess was part of the sensitive dataset (e.g. a patient's electronic health record). MIA, in fact, has a higher fidelity in graph-based settings, due to additional information that intrinsically lies in the structure of a graph [86]. Another commonly used attack is termed an attribute (or feature) inference attack [44]. It aims to reconstruct sensitive features of individuals in the training dataset and typically involves an adversary having access to a non-overlapping dataset of publicly available attributes which, alongside the predictions of the trained model, are used to determine the value of a sensitive feature that belongs to a target participant. Furthermore, models trained on graph-structured data, such as GNNs, were shown to be susceptible to model inversion attacks (MInv) [34], which allow the adversary to extract sensitive training data by leveraging the internal representations of the model (e.g. reverse-engineering a model update into disclosing which data point corresponds to this specific update). Authors in [132] show that MInv attacks can be adapted to graph-based learning. Wu et al. [120] introduce a privacy attack called *LinkTeller*, which recovers private edges in GNNs. Notably, seeing as graph-structured data captures information not just about individuals themselves, but about their relationships with other participants, all of these attacks can potentially compromise privacy of *multiple participants at once*.

The increasing popularity of graph-based analytics and machine learning coupled with the regulatory and ethical mandates to protect sensitive data imply that privacy enhancing technologies (PETs) [54] need to be applied in order to provide formal guarantees of privacy. Differential privacy (DP) [27] was proposed to objectively quantify the privacy loss of individuals whose data is subjected to algorithmic processing and is now regarded as the *gold standard* of formal privacy guarantees. Differentially private algorithms upper-bound the amount of information that can be inferred by an adversary who observes a computation's output, thus mitigating the attacks discussed above (*central* DP). The utilisation of DP mechanisms thus allows for e.g. training machine learning models on sensitive datasets while preserving the privacy of the contributors' data. Alternatively, DP can be applied directly to the data (*local* DP), allowing it to be publicly released for subsequent analytics or machine learning tasks. A third approach sees DP being used to generate *synthetic* data which shares statistical attributes with some population without the associated privacy risks.

However, the adaptation of any of the aforementioned concepts to graph-structured data is non-trivial for two main reasons: (1) there exist several notions of DP on graph-structured data, which protect different components of the graph, and thus need to be selected carefully and appropriately to the application; (2) due to the formal definition of DP, its realisation on graphs encompasses several additional challenges compared to tabular data.

To promote the development of responsible and privacy-preserving sensitive data processing systems, we identify the requirement for a comprehensive systematisation of knowledge. In this work, we investigate existing DP implementations, their limitations and application areas, as well as a number of challenges associated with differentially private learning on graph-based structures and promising directions for future work. We distinguish three lines of works: (1) non-machine learning graph analytics methods, (2) machine learning approaches on graph-structured data with graph neural networks (GNNs), and (3) generative models on graphs. This distinction allows us to emphasise open challenges and highlight opportunities to transferring DP techniques from graph analytics methods to GNNs. The outline of the remaining work and our main contributions can be summarised as follows:

- In Section 2, we provide an introduction to graph-structured data and graph neural networks, as well as a formal definition of DP;
- We formalise the three main notions of central DP on graph-structured data: *edge-level*, *node-level*, and *graph-level* DP in Sections 2 and expand them by introducing several additional notions of DP, including local DP, in Section 4;
- We demonstrate how different DP formulations can be applied in various settings in Section 5 and how graph analytics and graph learning under DP can be compared in these scenarios;
- We identify limitations and open challenges of these approaches and pinpoint promising areas of future work in the domain of DP on graph-structured data in Section 6.

2. Background

In this section, we formalise the concept of DP, introduce the three main notions of DP on graph-structured data, as well as the concept of (global) sensitivity, the Gaussian and the Laplace mechanisms, and provide a brief introduction to graph-structured data and graph neural networks (GNNs).

2.1. Graph-Structured Data.

In the following, we will refer to a graph $G = (V, E)$ as a collection containing a set of nodes $V = \{v_1, v_2, ..., v_n\}$ and a set of edges $E = \{e_1, e_2, ..., e_m\}$, n and $m \in \mathbb{N}$. Here, n determines the number of nodes in the graph and m the number of edges. The data contained in the graph can be split into the attributes contained in the nodes V of the graph, which can be referred to as node features, and the data held by the connections E between the nodes. The edges can optionally also contain edge attributes, holding additional information about the tightness or nature of the connection.

2.2. **Differential Privacy.** Differential privacy (DP) is a stability condition on randomised algorithms that makes their outputs approximately invariant to the inclusion or exclusion of a single individual [27]. In the words of the authors of [27], DP promises "to protect individuals from any additional harm that they might face due to their data being in the private database that they would not have faced had their data not been part of [the database]". This allows one to interpret DP as guaranteeing an upper bound on the *effect size* introduced by the inclusion or exclusion of the individual's data [110] on subsequent computations. The DP framework and its associated techniques allow data analysts to draw conclusions about datasets while preserving the privacy of individuals. We note that the DP guarantee makes *no assumption* about potential correlations between datapoints, however the "standard" interpretation of DP can behave in unpredictable ways when applied naïvely to data with such correlations such as graphs, prompting the more specific definitions introduced below. In the sequel, we will first discuss central DP; we outline its differences to local DP in Section 2.2.1 and 4.6.

In a setting of central DP on graph-structured data, we assume that an analyst $\mathcal{A}$ is entrusted with a database D containing sensitive graph-structured data. From D a neighbouring (in this work, we additionally use the term *adjacent*) dataset D' is constructed by either (a) removing or adding one node and its adjacent edges (*node-level* DP), (b) removing or adding one edge (*edge-level* DP), or (c) removing or adding one graph (*graph-level* DP). Although we will use this *add/remove* notion of adjacency in the definitions throughout this paper, we note that there also exist two other general adjacency notions: (1) *replacement* adjacency, where a sensitive record is replaced with another sensitive record and (2) *dummy replacement* adjacency, where a sensitive record is replaced with a dummy record. The latter is conceptually equivalent to the *add/remove* notion. We stress that it is important to formally state which adjacency notion is used when discussing DP, as the privacy guarantee changes under different adjacency notions. For graph-structured data on single-graph datasets, the *add/remove* notion is typically utilised, since the replacement of an edge or a node would affect more nodes in the graph and therefore increase the distance between the adjacent datasets. Besides the aforementioned, additional notions of adjacent datasets have been introduced. For example, Blocki et al. [8] define a notion of adjacent datasets for labelled graphs, where the change of a labelling function of a single node also results in an adjacent database. Other works [43, 59] also include the removal of an *isolated* node to a suitable method to gain adjacent databases.

Formally, DP can be defined as follows:

Definition 2.1 ((ε-δ)-DP)**.** *A randomised algorithm $\mathcal{M} : X \to R$, where X is a collection of sensitive databases, is $(\varepsilon$-$\delta)$-differentially private if, for all $S \subseteq R$ and all neighbouring databases D and D' in X, the following statement holds:*

$$\mathbb{P}[\mathcal{M}(D) \in S] \leq e^{\varepsilon}\mathbb{P}[\mathcal{M}(D') \in S] + \delta, \tag{2.1}$$

where the guarantee is given over the randomness is $\mathcal{M}$ and holds equally when D and D' are swapped. In the sequel, we will only describe one pair of neighbouring databases/graphs for brevity but stress that the statements must also hold when the neighbouring relationship is reversed.

The definition of neighbouring datasets on graph-structured data depends on the desired formulation of privacy in the setting (i.e. which attributes need to be kept private, such as outgoing edges for instance). Therefore, the desired notion (as well as the associated mechanisms) of privacy preservation depend on what the data owner requires to protect, the structure of the graph and the desired application to ensure a context-appropriate interpretation of the DP guarantee. In order to employ differentially private algorithms to process graph-structured data, the property of neighbouring datasets thus needs to be formally defined. The three main notions of DP on graphs can be formalised as follows:

Definition 2.2 (*Edge-level* DP)**.** Under *edge-level* DP, two graphs G and G' are neighbouring if they differ in a single edge (either through addition or removal of the edge) [55]. (ε-δ)-edge-DP is therefore preserved if equation (2.1) holds for all events S and all pairs of neighbours G, G' that differ in a single edge. In this setting, two graphs $G = \{V, E\}$ and $G' = \{V', E'\}$ are neighbours if

$$V' = V \wedge E' = E \setminus e_i, \tag{2.2}$$

where $e_i \in E$.

Zhu et al. [135] term edge-level DP in the setting of undirected graphs *degree*-DP.

Definition 2.3 (*Node-level* DP)**.** Under *node-level* DP, two graphs $G = \{V, E\}$ and $G' = \{V', E'\}$ are defined as neighbouring if they differ in a single node and its corresponding edges (achieved through a node removal/addition) [2]. (ε-δ)-node-DP is therefore preserved if equation (2.1) holds for all events S and all pairs of neighbours G, G', that differ in a single node and its corresponding edges:

$$V' = V \setminus v_i \wedge E' = E \setminus c, \tag{2.3}$$

where v_i is a node in V and c is the set of all edges connected to v_i.

Figure 1 visualises these two main definitions of DP on graphs. Two neighbouring datasets (graphs) under *node-level* DP and *edge-level* DP are displayed in sub-figures **A** and **B**, respectively.

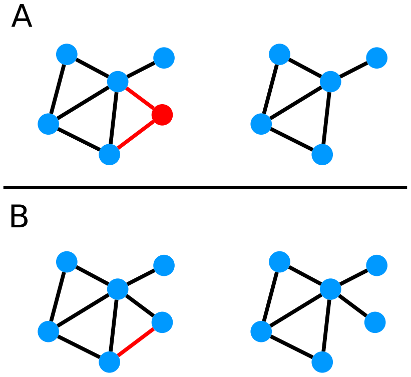

Figure 1: Two neighbouring graphs in the context of (**A**) *node-level* DP and (**B**) *edge-level* DP. By removing (**A**) one node and its adjacent edges or (**B**) one edge (displayed in red), two neighbouring graphs can be transformed into each other.

For multi-graph datasets, we can define a different notion of DP:

Definition 2.4 (*Graph-level* DP)**.** Under *graph-level* DP, we define two multi-graph datasets $D = \{G_{11}, G_{12}, \dots, G_{1n}\}$ and $D' = \{G_{21}, G_{22}, \dots, G_{2m}\}$ to be neighbours if they differ in one single graph (achieved through the addition or removal of one entire graph). (ε-δ)-graph-level DP is therefore preserved if equation (2.1) holds for all events S and all pairs of neighbouring datasets D and D', where

$$D' = D \setminus G_{1i}, \tag{2.4}$$

and $G_{1i} \in D$.

We now assume that the analyst $\mathcal{A}$ executes a function (or *query*) f over the graph dataset. When considering DP in GNNs, the function f is a repeated composition of the forward pass, loss calculation, and gradient computation of the graph neural network (resulting in a "database" of gradients). In order to determine the magnitude of noise that needs to be added, we are required to calculate the global sensitivity of the function that noise is applied to. We will consider either the global L_1- or the L_2-sensitivity of f.

Definition 2.5 (Global L_2-sensitivity Δ_2 of f)**.** Let f be defined as above and X be the set of all neighbouring databases. We can define the (global) L_2-sensitivity of f as:

$$\Delta_2(f) := \max_{D,D' \in X, D \simeq D'} \|f(D) - f(D'))\|_2. \tag{2.5}$$

We note that the maximum is taken over all neighbouring pairs of datasets in X.

Using the definition of L_2-sensitivity, we can formalise the Gaussian Mechanism on f:

Definition 2.6 (Gaussian Mechanism)**.** Let Δ_2 and f be defined as above. The Gaussian Mechanism $\mathcal{M}$ is applied to the function $\mathbf{y} = f(x)$, $y \in \mathbb{R}^n$, as follows:

$$\mathcal{M}(\mathbf{y}) = \mathbf{y} + \xi, \tag{2.6}$$

where $\xi \sim \mathcal{N}(0, \sigma\mathbb{I}^n)$. $\mathbb{I}^n$ is the identity matrix with n diagonal elements and σ is calibrated to Δ_2.

Similarly to L_2-sensitivity, we can define the L_1-sensitivity as:

Definition 2.7 (Global L_1-sensitivity Δ_1 of f)**.**

$$\Delta_1(f) := \max_{D,D' \in X, D \simeq D'} ||f(D) - f(D')||_1. \tag{2.7}$$

When it is clear from context, we will omit the argument and write just $\Delta_{1/2}$.

Definition 2.8 (Laplace Mechanism)**.** Let Δ_1 and f be defined as above. The Laplace Mechanism $\mathcal{M}$ is applied to the output $\mathbf{y} = f(x)$, $\mathbf{y} \in \mathbb{R}^k$, as follows:

$$\mathcal{M}(\mathbf{y}) = \mathbf{y} + (\xi_1, \xi_2, \dots, \xi_k), \tag{2.8}$$

where ξ_i are I.I.D. draws from Lap $\left(0, \frac{\Delta_1}{\varepsilon}\right)$.

2.2.1. *Local and Central DP.* As briefly mentioned above, one can furthermore distinguish between *local* and *central* DP. Under local differential privacy (LDP) [124] the data owner performs the noise perturbation step before the data reaches the analyst. Such interpretation can be preferable in low-trust collaborative learning settings, as no party other than its owner has access to the data before the learning task commences. Data owners only share a perturbed version of their data, which reduces the amount of information an analyst can infer about the shared data itself, while still allowing to draw insights from the privatised aggregated data [98]. Note that in local DP, adjacency is defined differently as in central DP (see Section 4.6). Local DP thus bounds the information at the data source itself, minimising the potential privacy exposure [54]. An adversary is, therefore, unable to infer the input value with high confidence, but is possible to approximate the target query if provided with a large number of noisy samples [98]. More details about local DP on graph-structured data can be found in Section 4.6.

When DP is, on the other hand, applied to the output of the computation instead of the input data, one speaks of central differential privacy. In this case, the noise is not added directly to the input data but instead to the computation outputs. Due to the properties of DP, only a bounded quantity of additional information can be derived about the data belonging to an individual, while the overall statistics of the whole dataset can still be approximately evaluated.

2.3. **Graph Neural Networks.**

To allow for machine learning to be performed directly on graph-structured data, GNNs were proposed [101]. They leverage the full underlying structure of the dataset and maximise learning capacity by directly learning *on the graph*. GNNs can be applied to either single graph or multi-graph datasets, depending on the desired task and available dataset. The three major tasks of GNNs are *node-level prediction* (where one label is predicted for each node in the graph), *edge-level prediction* (where edges are predicted or labeled), and *graph-level prediction* (where one label is predicted for each graph).

A key concept of GNNs is message passing [61], where information is shared along edges and therefore propagated among neighbourhoods of nodes. This property enables the utilisation of the full dimensionality of graph datasets. However, this typically complicates the disentanglement of contributions by individual nodes, making the calculation of individual privacy loss per each participant a challenging task.

3. Systematisation Methodology

We conducted a survey of papers that intersect the domains of graph analytics, deep learning on graphs, or graph generation with DP. We employed the *Google Scholar* and the *Web of Science* search engines and examined papers that contained the keywords “node-”, “edge-”, “graph-” “differential privacy” between January, 2007 and February 2023. Our searches often had to be coupled (e.g. “node differential privacy graphs”), as notions such as *graphs* or *nodes* are often used in unrelated concepts such as computation graphs or network nodes. We selected 57 studies, which we partitioned based on the DP formulations employed in each work: *node-level* DP, *edge-level* DP, *graph-level* DP, and whether *local DP* was applied in the respective works. Furthermore, we separated the works into *graph analytics* and *GNN training.* We additionally recorded the contexts in which DP was applied. A summary of the works that we discuss in this study can be found in Tables 1 and 2.

We observed that a large number of studies concentrate on the usage of graph datasets but explicitly not on the utilisation of GNNs. The large amount of research in the context of DP on graphs in general shows the importance of applying differentially private algorithms to graph-structured data. However, applications of DP to GNNs are currently underrepresented, presumably due to the fact that GNNs are a relatively recent deep learning method, and the application of DP to GNNs entails several challenges. For example, there is no singular explicit notion of "DP" in different graph machine learning settings, as discussed below. We are optimistic that the here-presented systematisation of different possibilities to apply differential privacy to graph neural networks can act as a guide to practitioners and aid them in the the development of new methods in this area. With the advent of privacy-preserving machine learning and the strong interest in geometric deep learning applications, we strongly believe the differentially private training of GNNs to be a promising future research area with several applications to sensitive data. We therefore explicitly decided to include both graph analytics and machine learning on graphs in our survey. Some exemplary application areas are discussed in Section 5.

4. DP formulations on Graph-Structured Data

In this section, we outline and discuss methods from the research field of differentially private graph analytics and graph machine learning. We identify and consider three separate lines of work: **(a)** DP in statistical graph analytics methods, **(b)** in graph neural networks (Table 1), and **(c)** in generative models (Table 2). We also indicate the notion of DP that was applied in the respective research in the columns *Edge-DP*, *Node-DP*, *Graph-DP*, and *LDP* and summarise ranges of the privacy budget ε if they were reported in the respective works. The line of work of DP in traditional graph analytics **(a)** includes e.g. methods for privately computing graph statistics like degree-distributions [43], frequent sub-graph-mining [103], and sub-graph counting [8]. The works of DP for GNN training **(b)** include, for instance, text classification [47], whole-graph classification [81], and attacks on GNNs [132] and the works on DP graph generation **(c)** are summarised in Table 2.

The first differentially private computation on graph data was introduced by Nissim et al. [84]. The authors implement a DP algorithm to estimate the computational cost of minimum spanning tree creation and triangle counting. In their work, the authors opted for the utilisation of *edge-level* DP.

As indicated in Tables 1 and 2, we generally observe a focus on *edge-level* DP in earlier papers, compared to a more frequent utilisation of *node-level* DP in more recent works. We attribute this to the fact that *node-level* DP is more challenging to achieve, but offers stronger privacy guarantees (as it considers the privacy of a node and all its adjacent edges). Works on *graph-level* DP are quite rare. However, we believe this notion of DP to be promising and given that different works name the same concept differently, we still included graph-level DP in Tables 1 and 2.

We furthermore observe that, in the works discussing DP on GNNs, authors frequently omit to specifically assign the guarantees provided to one of the aforementioned DP notions, which highlights the need for more systematic approaches to defining DP in graph learning tasks. We attribute this lack of specification to missing systematisation of terminology in this area as well as the challenging task to differentiate the individual notions of privacy in graph learning tasks and their dependence on the dataset and the application area.

	Edge-DP	**Node-DP**	**Graph-DP**	**LDP**	**Year**	**Reference**	**Context**	ε
Statistical Graph Analytics	✓	✓			2007	Nissim et al. [84]	Estimation for spanning trees	-
	✓				2009	Hay et al. [43]	Graph degree estimation	$[0.01; 1]$
					2011	Gehrke et al.** [37]	Zero-knowledge statistics estimation	-
	✓				2011	Machanavajjhala et al. [73]	Privacy in social graphs	$[0.5; 3]$
	✓				2011	Sala et al. [100]	Release of private graphs	$[0.1; 100]$
	✓				2011	Karwa et al. [55]	Private subgraph counting	0.5
	✓				2012	Gupta et al. [40]	Private cut function release	-
	✓				2012	Karwa et al. [56]	Release of graph degree sequences	-
	✓				2012	Mir et al. [78]	Private release of graph distribution	0.2
	✓	✓			2013	Blocki et al. [8]	Restricted sensitivity for DP	-
		✓			2013	Chen et al. [14]	Private graph database aggregation	$[0.1; 0.5]$
		✓			2013	Kasiviswanathan et al. [59]	Private graph analysis	-
			✓		2013	Shen et al. [103]	Private graph pattern mining	$[0.1; 1]$
	✓				2013	Wang et al. [117]	Private spectral graph analysis	460
	✓				2013	Wang et al. [115]	Private spectral graph analysis	-
	✓				2014	Chen et al. [13]	Correlated network data release	$[0.6; 1]$
	✓				2014	Lu et al. [72]	Estimation of graph model parameters	$[0.1, 1]$
		✓			2014	Raskhodnikova et al. [94]	DP analysis of graphs	-
	✓	✓	✓		2014	Task et al. [108]	Private social network analysis	-
		✓			2016	Day et al. [21]	Private graph distribution release	$[0.1; 2]$
	✓				2016	Jorgensen et al. [51]	Private attributed graph models	$[1; 20]$
		✓			2016	Raskhodnikova et al. [95]	Private release of graph statistics	-
	✓			✓	2016	Wang et al. [116]	Private aggregation of data	$[0; 2]$
	✓	✓			2017	Zhu et al. [136]	Applications of differential privacy	-
	✓	✓		✓	2018	Cormode et al. [17]	Private data release	-
		✓			2018	Macwan et al. [74]	Private release of graph data	0.5
	✓				2019	Arora et al. [3]	Graph sparsification	-
		✓			2019	Sealfon et al. [112]	Estimation of graph statistics	-
					2019	Sun et al. [107]	Subgraph statistics, decentralised DP	$[1; 10]$
		✓			2019	Yuxuan et al. [128]	Private histogram release	-
	✓				2020	Chen et al. [15]	Private synthetic data release	$[2; 5]$
		✓			2020	Liu et al. [70]	Node strength distribution	$[0.1; 2]$
		✓			2020	Zhang et al. [131]	Private social graph release	$[0.1; 20]$
		✓		✓	2020	Zhang et al. [129]	Control-flow graph coverage analysis	$[2^{-5}; 2^5]$
		✓			2021	Iftikhar et al. [46]	Private release of degree distribution	$[0.01; 10]$
		✓			2021	Fichtenberger et al. [32]	Private dynamic graph algorithms	-
	✓			✓	2021	Imola et al. [48]	Private sub-graph counting	$[0; 2]$
		✓			2021	Lan et al. [63]	Private node strength histogram release	$[0.1; 2]$
		✓			2021	Liu et al. [71]	Private degree histogram release	$[0.1; 2]$
		✓			2021	Sealfon et al. [111]	Private graph density estimation	-
	✓	✓		✓	2021	Xia et al. [122]	Benchmark platform for DP on graphs	-
	✓			✓	2021	Zheng et al. [134]	Private graph publication framework	-
GNNs				✓	2020	Sajadmanesh et al.* [98]	Locally private GNNs	$[0.01; 3]$
		✓			2021	Daigavane et al. [20]	Node-level DP in GNNs	$[5; 30]$
					2021	Igamberdiev et al.* [47]	Private text classification	$[1; 100]$
					2021	Olatunji et al.* [85]	Private GNN and graph data release	$[1; 40]$
					2021	Zhang et al.* [132]	Attacks on GNNs	$[1; 10]$
			✓		2022	Mueller et al. [81]	Graph-level DP for graph classification	$[0.5; 20]$
	✓	✓			2022	Sajadmanesh et al. [99]	Aggregation perturbation	$[0; 16]$

Table 1: Summary of existing works addressing DP statistical and learning-based analytics on graphs, in ascending order by publication year and alphabetically within the same year. Ticks in columns **Edge-DP**, **Node-DP**, and **Graph-DP** specify which notion of privacy was used. A tick in column *LDP* indicates that the authors used local DP. One asterisk (*) indicates that the DP notion is not clearly stated. Two asterisks (**) the utilisation of zero-knowledge privacy (see Section 4.5.3). The column ε reports the evaluated privacy budgets in the respective works.

4.1. **Sensitivity Calculation on Graphs.** As described above, ensuring data privacy on graphs presents additional challenges compared to structured databases such as image or tabular datasets, since the data points are inter-connected and the graph structure itself can contain sensitive information. Furthermore, depending on the application it can be desired

	Edge-DP	**Node-DP**	**Graph-DP**	**LDP**	**Year**	**Reference**	**Context**	ε
Generative	✓				2009	Mir et al. [79]	Synthetic graph generation	-
		✓			2014	Proserpio et al. [91]	Synthetic graph generation	$[0.01; 10]$
		✓			2015	Borgs et al. [9]	Graphon estimation	-
	✓				2016	Karwa et al. [57]	Synthetic graph generation	-
	✓	✓		✓	2017	Qin et al. [92]	Graph generation for social networks	$[0; 7]$
		✓			2018	Borgs et al. [10]	Graphon estimation	-
	✓				2019	Zhu et al. [135]	Graph generation for social networks	[0.5; 3]
	✓				2021	Zheng et al. [133]	Network Generation	$[0.1; 440]$

Table 2: Summary of existing works on DP graph generation methods, in ascending order by publication year and alphabetically within the same year. Ticks in columns **Edge-DP**, **Node-DP**, and **Graph-DP** specify which notion of privacy was used. A tick in column *LDP* indicates that the authors used local DP. The column ε reports the evaluated privacy budgets in the respective works.

to protect different parts of the graph. One fundamental challenge is therefore the issue of sensitivity calculation.

In cases of graphs, this value can be challenging to obtain as it depends not only on the structure of the graph but also on the attributes of the query function. Two main methods have been proposed to obtain node differentially private algorithms which are either based on (a) the utilisation of projections, for which sensitivity can be bounded, or (b) on computing Lipschitz extensions [8, 14, 95]. Raskhodnikova et al. [95] study the efficient computation of Lipschitz extensions for multi-dimensional functions on graphs, which can be obtained in polynomial time, and determine that they do not always exist - in comparison to Lipschitz extensions for one-dimensional functions. Karwa et al. [55] discuss methods to determine the local and smooth sensitivity of DP graph analysis for example for triangle count, k-star estimation, and sub-graph counting queries. As noted above, the adjacency notion also influences the sensitivity calculation, as sensitivity is computed over all adjacent database pairs. Moreover, a small portion of the aforementioned papers do not utilise the global sensitivity, but a relaxation such as local sensitivity. For details on local sensitivity and other variations, we refer to [23]. Unless otherwise indicated, we will limit ourselves to global sensitivity in this work.

In the next sections, we give more details about the different definitions of DP on graphs in *node-level*, *edge-level*, *graph-level* DP as well as some alterations and combinations of these, with respective interpretations of what is implied by neighbouring datasets in each setting.

4.2. **Edge-Level Differential Privacy.** There exist several approaches that allow one to release graph statistics with *edge-level* DP guarantees, including sub-graph counts [56], spanning tree estimation [84], degree distributions [43, 108] and graph cuts [40]. Those settings set a focus on privatising the relationships between nodes. This can be applied to social network graphs [43, 79] or location graphs [123], where the edges contain sensitive information, but the data represented in the nodes of the graph are assumed to be publicly known or non-sensitive.

4.3. **Node-Level Differential Privacy.** *Node-level* differential privacy is a strictly stronger guarantee than edge-level differential privacy [59]. This is of particular importance in scenarios where graphs are very sparse, and thus, the removal of a single node can alter the graph structure severely. For instance, the number of triangles in a graph with n nodes can increase by $\binom{n}{2}$ when inserting a single additional node. Consequently, these functions tend to have high sensitivity [95], resulting in a large noise magnitude. Bounded-degree graphs (graphs where each node has an upper limit of edges and the degree of each node is therefore bounded) can assist in lowering the sensitivity. Here, the removal of a single node results in an upper-bounded change in edges which typically leads to a reduced impact on the output of the algorithm. When calculating the number of triangles in a graph, for instance, maximum change of a D-bounded-degree graph is $\binom{D}{2}$ which is strictly smaller than $\binom{n}{2}$ if $D < n$.

Settings that can benefit most from this formulations of DP are those that put an emphasis on the data within the node itself yet additionally privatise the connections between the nodes. This includes studies on social networks [8, 92], the publication of higher-order network statistics [46, 71, 74], and recommendation systems [73].

4.4. **Graph-Level Differential Privacy.** So far, *graph-level* DP has not been explored in great detail, neither in the context of graph analytics nor in GNNs. Task et al. [108] name this notion of privacy *partition privacy* and show its application to graph analytics of social networks. Shen et al. [103] investigate the mining of frequent graph patterns in multi-graph datasets and apply the mechanism of *graph-level* DP to their algorithm. They use Markov Chain Monte Carlo (MCMC) random walks to discover frequently appearing sub-graphs in the graph dataset and infer graph statistics under graph-level DP.

In the context of GNN training, *graph-level* DP can be applied in learning settings that investigate graph classification tasks, e.g. drug discovery or molecule classification [25], discovering disease-specific biomarkers of brain connectivity [65, 67], or shape analysis [119]. This way, privacy guarantees can be given to the individuals, whose sensitive information is contained in those multi-graph datasets. For instance, in the setting of drug discovery, a group of pharmaceutical companies can collaborate on a graph classification task, while bounding the information that can be inferred about their individual molecules, which represent the private data in this context. Mueller et al. [81] apply graph-level DP for classification tasks on several sensitive datasets, implementing the concept of graph-level DP on GNNs and showing potential applications.

4.5. **Further Definitions of DP on Graphs.** We consider node-, edge-, and graph-level DP to be the three main categories of DP guarantees on graph-structured data. However, there exist additional notions of DP that have not yet found a widespread application and are mostly derived from the notions formalised above. Here, we provide further details about those additional definitions and variations of applied notions of DP.

4.5.1. *k-Edge Differential Privacy.* One such formulation is *k-edge differential privacy* introduced by Hay et al. [43]. It defines a stricter notion of *edge-level* DP, where two graphs $G = \{V, E\}$ and $G' = \{V', E'\}$ are neighbours if $|V \oplus V'| + |E \oplus E'| \leq k$. Hereby, $\oplus$ denotes the symmetric difference. If $k = 1$, the definition recovers *edge-level* DP. However, if $k = |V|$, *k-edge-level* DP is a stricter definition than *node-level* DP, as the set of neighbouring graphs in the definition of *node-level* DP is a subset of the neighbouring graphs under k-edge-level DP. For nodes with a degree smaller then k, k-edge-level DP provides an equivalent protection as *node-level* DP. Nodes with a degree $\geq k$ face more exposure, since they have more edges.

However, one can argue that those high degree nodes have a higher impact on the general graph structure and it might therefore be necessary to expose them to larger privacy risks to allow analysts to accurately measure graph statistics. The authors experimentally evaluate their notion of k-edge-differential privacy on social network data from Flickr, LiveJournal, Orkut, and YouTube.

4.5.2. *Out-Link Differential Privacy.* Another definition of DP on graphs was introduced by Task et al. [108] and is termed *out-link differential privacy.* In this context directed graphs are considered, where it is possible to distinguish between incoming and outgoing edges of nodes. Under this notion, two datasets are considered to be neighbouring if all *out-links* (outgoing edges) of an arbitrary node are added or removed. Formally, two graphs $G = \{V, E\}$ and $G' = \{V', E'\}$ are neighbours, if $V = V'$ and $E' = E - \{(v_1, v_2) | v_1 = x\}$ for an $x \in V$. (v_1, v_2) hereby defines an edge going from node v_1 to node v_2.

Out-link DP is strictly weaker then *node-level* DP, but in many scenarios comparable to *edge-level* DP. Under this notion of DP, an attacker would not be able to determine whether a person x contributed their data to the construction of the graph and participants in the graph can hide their out-links. In the setting of a social network, for instance, a person x can deny friendships. Others can still claim to be friends with person x, but the latter can deny that those connections are mutual (i.e. that person x has out-going links to adjacent nodes). The authors argue that *out-link* privacy simplifies sensitivity computation and reduces noise addition requirements, enabling queries that would be infeasible under previous DP definitions.

Similar to k-edge-level DP, *out-link* DP can also be extended to k-out-link privacy. In this case, neighbouring datasets are considered, that differ in k out-links compared to the original dataset. When considering 2-out-link privacy, for example, two nodes can simultaneously deny all their out-links. This would also enable to protect a complete mutual edge, resulting in *edge-level* DP in addition to *out-link* DP.

4.5.3. *Zero-Knowledge Privacy.* Gehrke et al. [37] introduce a stricter formulation of *node-level* DP, namely *zero-knowledge privacy* on graphs, which authors argue is particularly desirable in social network analysis. It relies on a notion similar to the one of cryptographic zero-knowledge proofs [33], which entails that a protocol participant obtains a computation result with "zero additional knowledge" about the data used to perform this computation. A privacy mechanism $\mathcal{M}$ is ($\mathbf{Agg}$, ε)-zero-knowledge private if there exists a simulator $\mathcal{S}$ and an *agg* from the family of algorithms $\mathbf{Agg}$ such that for all neighbouring datasets D_1 and D_2 the following holds: $\mathcal{M}(D_1) \approx_\varepsilon \mathcal{S}(\mathbf{Agg}(D_2))$ [37]. Authors in [37] apply this definition to ensure that a mechanism does not release additional information apart from "aggregate information" which is considered acceptable to release to ensure usability.

4.5.4. *Relationship Differential Privacy.* Imola et al. [48] introduce a notion called *relationship DP*, a definition falling under local DP. Here, one edge in a graph is masked during the entire learning process. In a setting of social network analysis, relationship DP assumes that each user only knows their own connections (i.e. friends), requiring users to have a higher degree of "trust" when interacting with their immediate neighbours. Given two users v_i and v_j that share a link in the social network, under relationship-DP a user v_i has to trust its adjacent user v_j not to leak information about their shared connection. Intuitively, *edge-level*

LDP considers the edge from user v_i to user v_j and the edge from user v_j to user v_i to be two separate "secrets", whereas relationship DP assumes that the two edges represent the same "secret". (More details about edge-level LDP can be found in Section 4.6.) Therefore, the trust model of relationship DP is a stronger one than the one of *edge-level LDP*, which does not hold any assumptions about what other users do, but weaker than the one of centralised edge-level DP, where all edges are held by a centralised party. If a randomised algorithm $\mathcal{M}$ provides ε-edge-level LDP, then $\mathcal{M}$ provides 2ε-relationship DP, given that an edge (v_i, v_j) affects two elements in the adjacency matrix of the graph and the property of group privacy [27].

The authors apply this formulation of privacy to algorithms for sub-graph, k-star, and triangle counting, which can be used to analyse connection patterns in graphs.

4.5.5. *Edge-Weight Privacy.* For shortest path or distance queries on graphs, edge-level and node-level DP are not well suited, since both queries usually return a set of edges, which violates both edge-level and node-level DP. Therefore, Sealfon [102] introduced a different notion of privacy on graphs: *edge-weight privacy.* This notion of privacy is applicable if the edge weights of a graph contain private data, whereas the graph structure itself is publicly available and does not need to be protected. An example would be traffic data in a known street system.

4.5.6. *Node Attribute Privacy.* Chen et al. [15] define another notion of privacy for attributed graphs. An attributed graph $G = (V, E, A)$ is the set of vertices V, edges E and node attributes A. In this definition of privacy, two graphs are defined to be neighbouring if they differ in one edge or in the attribute vector of one node. So in this scenario, the presence of nodes is assumed to be non-private, whereas the connections (edges) between the nodes as well as the attributes that define the nodes contain private information. This definition can for example be useful in social networks, where the existence of a profile can be publicly known but friendships and personal attributes (stored in the profiles/nodes) are private.

4.6. **Local DP on Graphs.** There exist several works that target the preservation of local differential privacy (LDP) on graph-structured data. The advantage of local DP [58] in comparison to central DP is that no trusted third party is required. LDP can and has been applied to both classical graph analytics and graph neural networks. Qin et al. [92] define *edge-level* and *node-level* LDP in the context of neighbour lists. A neighbour list of a vertex v_i in a directed graph with n vertices is defined to be an n-dimensional bit vector $(b_1, \ldots, b_n)$, where $b_i = 1$, $i \in [1; n]$, if and only if there exists an edge (v_i, v_j), going from v_i to v_j, in the graph, otherwise $b_i = 0$. *Edge-level LDP* is then defined for two neighbour lists that differ in exactly one bit, whereas *node-level LDP* is defined for any two neighbour lists.

4.6.1. *Locally private graph analytics.* Examples for LDP in graph analytics tasks include Zhang et al. [129], who perform control-flow graph coverage analysis under *node-level* LDP and Imola et al. [48], who apply LDP to sub-graph counting, k-star and triangle counts while preserving *edge-level* LDP.

4.6.2. *Locally private GNNs.* LDP can also be applied to GNNs, where settings such as decentralised social networks can benefit from this property, as shown by Sajadmanesh and Gatica-Perez [98]. They introduce a privacy-preserving architecture-agnostic GNN algorithm, which preserves private node features under LDP. Their architecture includes an LDP encoder and an unbiased rectifier, which functions as the communicator between the server and the graph. This algorithm can be applied in a setting where either the node features or the labels (or in certain cases both) are to be kept private regardless of the GNN architecture. Authors use a so-called *multi-bit mechanism* which allows the nodes to perturb their features before passing them to the server. The server then processes this noisy data through the first convolutional layer. GNNs aggregate the node features before passing them through the activation function, which can be used as a denoising mechanism to average out the noise that was injected into the node features in the first place. The authors employ a generalised randomised response mechanism [52] to preserve privacy of node labels. However, they explicitly do not preserve *node-level* or *edge-level* DP but protect the privacy of node features and labels. This leaves the graph structure itself unprotected, which remains an open challenge in this context. Note that *label* DP is a distinct area of study which we do not cover here, but has been described for non-graph datasets e.g. in [29, 38, 75].

4.7. **DP for Graph Neural Networks.** While the notion of DP on traditional graph analytics and statistics applications (particularly for private data release) is well established, there exist significantly fewer studies on differentially private GNN training. This can be attributed to multiple factors, one of them being the number of different GNN machine learning settings (e.g. single- and multi-graph settings). This renders the identification of a standardised method for differentially private GNN training significantly more challenging. Furthermore, GNN learning is not yet a fully established area of research, leaving a number of learning contexts unexplored. In this section we introduce two methods that have been used to achieve differentially private training on GNNs.

4.7.1. *DP-SGD Training of GNNs.* One of the most common methods to perform differentially private training in (non-graph) machine learning is differentially private stochastic gradient descent (DP-SGD) [1]. Here, a gradient descent step is privatised through bounding the gradient L_2-norm (clipping) and through the addition of calibrated noise, such that the output of the gradient calculation over two neighbouring datasets can –with high probability– not be well distinguished. This concept is not limited to SGD and can be applied to other first-order optimisation techniques, e.g. Adam. In standard machine learning, the clipping in DP-SGD is applied to the gradient *of each individual data point* to minimise the amount of noise that has to be added to the gradients. This method, naturally befitting structured databases with well-defined notions of what an "individual" gradient entails, does not seamlessly extend to graph machine learning in all cases. For graph classification tasks, for instance, each graph can be seen as an individual entity in a multi-graph dataset and, therefore, *graph-level* DP can be seen as a natural formulation in these learning settings. Here the standard procedure of DP-SGD can be transferred from database queries to graph learning tasks, matching database entries (rows) with individual graphs. This has been shown in [81]. Even though graph-level DP has not been deeply explored in research so far, we believe this to be an interesting and promising research area holding multiple applications, for example in medical settings with population graphs or brain networks (see Sections 5.2 and 5.3).

DP-SGD was initially not designed with graph databases in mind, where connections and interactions between nodes complicate some fundamental aspects of DP-SGD, like per-sample gradients. In a graph, single data points (nodes or edges) cannot be separated from the whole dataset without breaking up the graph structure, which is essential to the message passing mechanisms of GNNs. Therefore, DP-SGD is not directly transferable to GNNs for node-level and edge-level DP on single-graph datasets. This not only precludes a notion of "per-sample" gradients, but also *privacy amplification by sub-sampling*, which states that a DP mechanism executed on a random (secret) sub-sample of a population results in tighter privacy guarantees than when applied to the whole population [4]. For DP relaxations like Rényi DP [80, 114] or Gaussian DP [24], different (sometimes stronger) amplification results hold.

When working with single-graph datasets, randomly sampled nodes are likely to result in disconnected nodes that do not function as a suitable sub-graph for learning. Therefore, appropriate sub-sampling techniques for using DP-SGD with GNNs have been developed. Igamberdiev et al. [47] implement a graph splitting method, which partitions the graph into smaller batches to approximate sub-sampling amplification and apply DP mechanisms to graph neural networks. Daigavane et al. [20] recently introduced a set of techniques to enable the training of node-differentially private multi-layer GNNs, whereas previous works were constrained to single-layer GNNs. They implement (sub-sampled) DP-SGD by sampling the local *neighbourhood* of a node and by analysing the of affected per-sample gradient terms in a sub-graph. So far however, no universal method to analyse privacy amplification by sub-sampling in GNNs on single-graph datasets has been proposed.

Of note, sub-sampling is not the only method to amplify privacy. It can also be amplified by shuffling [16, 28] and diffusion/post-processing [5]. However, to the best of our knowledge, no works have discussed these methods for graph-structured data. Given the so far limited amount of work in areas of DP-SGD training of GNN and amplification methods, we believe those to be important research questions, which need to be explored in more detail.

4.7.2. *Private Aggregation of Teacher Ensembles.* Differentially private stochastic gradient descent is one of the most common methods to offer DP guarantees in machine learning. However, there are also alternative methods of preserving DP in machine learning, one being private aggregation of teacher ensembles (PATE), introduced by Papernot et al. [87]. PATE and its variants (e.g. [50, 88, 126, 130]) leverage an ensemble (a collection) of so-called *teacher models* that are trained on disjoint datasets containing sensitive data. These models are not published but instead used as teacher models for a separate student model. The student model cannot access any single teacher model nor the underlying data. It instead relies on a noisy voting algorithm performed across all teacher models to make a prediction [87]. One notable limitation of PATE is the reliance on a publicly available unlabelled dataset that is utilised by the teacher model. In general, this is a rather strong assumption, particularly in contexts relying on scarce, private datasets, such as medical data, limiting how generally it can be adopted as the means of differentially private training. In general, PATE should be considered as a private student-teacher data labeling mechanism rather than necessarily representing a method for private collaborative training.

This shortcomings of PATE techniques are compounded by a low utility of PATE in graph settings as well as the limited generality beyond graph classification settings, as the physical separation of datasets in graph learning destroys structural information, significantly reducing the utility of the trained model [85]. Therefore, Olatunji et al. [85] recently introduced a framework named *PrivGNN*, which also leverages a student-teacher training paradigm for GNNs. The authors generate pseudolabels for public query nodes using specialised GNN models while adding noise to the predictions. The method requires two datasets: labeled private data for the teacher model and unlabeled public data for the student model. In the end, the public student model is released. It is trained using the noisy pseudo-labels and is differentially private based on the post-processing property of DP. The authors therefore implement a method for private release of trained graph neural networks and show their results on three node classification datasets.

4.8. **Generative Models and Synthetic Graphs.** The ability to generate synthetic data samples allows one to augment existing datasets with additional data points in a privacy-neutral way, resulting in more diverse data representations. We summarised these strands of work in Table 2. Synthetically generated graphs can improve the utility of the model trained on this data as well as empirically reduces the effectiveness of inference attacks [90]. There exist several works in the area [9, 10, 15, 56, 57, 79, 92, 133] that allow one to generate graph-structured data in a private manner. Chen et al. [15], e.g., explore synthetic graph generation of social graphs under *edge-level* DP and Qin et al. [92], e.g., resort to LDP to generate synthetic decentralised social graphs. In [56] and [57], Karwa et al. introduce DP β-models that can be used for synthetic graph generation - also under edge-level DP and Zheng et al. [133] resort to generative adversarial networks for their graph creation method. Borgs et al. [9, 10] rely on non-parametric models, which do not require a previous estimation of the parameters, and utilise graphons to generate synthetic graphs under node-level privacy guarantees.

Even though some works have already investigated synthetic graph generation, a number of limitations still hold. For example, privacy-utility trade-offs are much more profound in graph generation tasks, forcing the model owner to either deteriorate the privacy guarantees or to generate graphs of much lower utility. Furthermore, synthetic graph generation methods have not yet reached a high variety of application areas and are often still applied only to benchmark datasets. Since the task of private graph generation has so far not been widely investigated in real world settings, we identify this to be a promising area of future work in the graph domain.

5. Application Areas for DP on Graphs

In this section we discuss how our findings from above can be and have previously been applied to graph learning tasks in order to establish which formulations of DP are most suitable for each context, and give insights into a selection of potential application areas for DP on graph-structured data. Lastly, we provide an outlook on promising future research in those settings. We chose three exemplary learning contexts to allow us to cover all commonly used formulations of DP on graphs (i.e. node-level, edge-level and graph-level DP). Overall, more contexts relying on sensitive (or proprietary) data can benefit from a formalisation of DP, such as drug discovery [49] or location-based learning [60]. We leave an in-depth investigation of privacy in these settings as future work.

5.1. **Social Networks.** One of the more well-researched areas of private learning on graphs concerns social graphs [8, 37, 73, 74, 78, 92, 99, 100, 108], where the personally-identifying information is contained in the nodes of the graph and/or in the edges, defining the interactions between individuals, that could potentially allow to uniquely identify them (e.g. when spatio-temporal data is published [22]). As a result, there exist two concievable routes to perform private learning on such data: *edge-level* DP to protect the connections to other individuals in the graph and prevent unique identification of users like in [55, 73, 78, 92, 100] and *node-level* DP to protect the data of each individual itself (as well as the outgoing edges) like in [8, 37, 70, 131]. Furthermore, Sajadmanesh et al. [98] utilise locally differentially private GNNs in the context of social networks. The focus on social network data for the utilisation of DP in GNNs shows the high importance of protecting privacy in these settings, as well as the associated risks inherent to working with such datasets. Recently, the same authors [99] introduced a method that guarantees both node-level and edge-level DP.

5.2. **Population Graphs.** The large amount of medical data collected by multiple medical institutions as well as personally through wearable devices, for instance, lead to mounting challenges of structuring these multi-modal datasets. One approach of handling this data heterogeneity is the construction of population graphs, which have found widespread adoption in medical research [6, 41, 83]. These data structures allow to encapsulate the information about patients across multiple departments and time periods (e.g. spatio-temporal patient data [69]), leveraging much more relevant information and leading to better predictions. One such scenario could involve representing each patient as a node and the whole patient population/cohort by a graph comprising the individuals, as described e.g. in [35, 76]. Connections between patients can, for instance, be based on their similarity (like in [89]). An advantage of creating such patient population graphs for the implementation of DP mechanisms is that the graph can be explicitly degree-bounded, limiting the impact of individual nodes on the graph structure.

Alternatively, each node can be patient-specific data about a single individual collected at different times by various specialists. Either of these contexts would benefit from the utilisation of *node-level* DP in order to quantify and limit the amount of information revealed when node-level data is processed or released, as they are relying on extremely sensitive data contained in each node.

5.3. **Brain Networks.** Here, we give an example for a graph classification problem on multi-graph datasets. In such setting it is not the information contained in a single node or inter-node connections that need to be kept private, but rather the information contained in a graph as a whole. One prime example of such dataset that contains sensitive information on a whole-graph level, rather than on the level of its individual constituents is a brain network graph [12]. Such data is used extensively in neuroimaging problems [82, 105, 127]. However, similarly to most medical datasets, due to the difficulty of obtaining such data (both because of the complexity of the task as well as of the privacy concerns) it is essential that the learning task is augmented with a suitable privacy-preservation mechanisms. In the case of brain network graphs, information about the value of individual voxels, or single connections to other voxels in the brain network are not necessarily personally identifying. Nonetheless, a collection of such interconnected points is considered to be a particularly sensitive medical dataset and it thus needs to be protected. For this setting, *graph-level*

DP is a particularly suitable technique for data release. To date, there only exists a small number of such implementations of differentially private multi-graph learning [81] and we envision that such formulation can gain significance as part of the future work in the area. We recall that DP deep learning on brain graphs (with learning tasks similar to [97] for instance) can be implemented through a straightforward utilisation of DP-SGD, similarly to Euclidean contexts.

6. Challenges and Outlook

In this section, we discuss a number of challenges associated with differentially private graph analytics, some of which can be attributed to the inter-connected nature of graphs, while others are inherent to DP itself. Note that we also discuss a number of potential complications arising in DP GNN training like privacy accounting, privacy-utility trade-offs, computational performance and interpretability of DP on graph-structured data.

6.1. **Privacy Accounting.** Typically, in differentially private machine learning settings privacy loss can be bounded per individual data point (i.e. per image or table record), thus considering data points independently from each other, simplifying privacy loss accounting. However, due to the intrinsic inter-dependency of nodes in a graph, independence cannot be guaranteed and therefore quantifying the contribution of each individual becomes non-trivial. Thus, there arises a need for concrete definitions which would allow the data owner(s) to determine the exact formulation of differentially private training that is applicable in the specific application areas. As noted by [2], the guarantees given by *edge-level* DP and *node-level* DP have different implications, which are based on the exact features data owners wish to protect.

DP is inherently compositional, that is, DP algorithms composed with each-other yet again yield a DP algorithm [27]. However, the heterogeneous composition of different formulations of DP in a graph setting (e.g. simultaneously accounting for learning the adjacency matrix of a graph and for node classification) has not been studied previously, and we consider it a promising avenue for future research.

Beyond the privacy accounting techniques discussed in Section 4.7 above, a line of work has introduced techniques aiming to account for *personalised/individual* or *per-instance* privacy loss. Personalised DP (pDP) was introduced by Ghosh and Roth in 2011 [39] and later extended by Feldman et al. [31] and termed *individual privacy.* Personalised DP (pDP) captures a level of privacy for each individual in a dataset, considering the addition or removal of a data point from all possible datasets. In contrast, per-instance DP –introduced by Wang et al. [118]– considers the addition or removal of a data point from a fixed dataset.

The general aim of individual/personalised/per-instance DP is to give a "bespoke" privacy guarantee to each individual participating in a computation, typically combined with a method to automatically terminate their participation when their individual privacy budget is exhausted. As the process of deciding to continue or halt a computation by considering the currently spent privacy budget is an instance of *fully adaptive* composition, additional mechanisms are introduced: a privacy odometer (which tracks the privacy expenditure in the process of computation, without having to specify a privacy budget in advance) and the privacy filter (which stops the computation once the privacy budget is exceeded). The combination of these tools allows for a finer-grained control of the information that can be learned from each individual data point and potentially higher utility. The ability to compute the individual privacy loss can allow a selective removal of individual nodes (and their corresponding edges), resulting in a much finer control of individual privacy expenditure. This method can permit tighter privacy bounding in settings where amplification by subsampling is not possible. However, it is also limited in applicability whenever the notion of a single individual within the graph is ill-defined. According to the author of [118], per-instance DP can be extended to graph-structured data in a straightforward fashion, but no published literature exists on the topic as of the writing of this paper. Of note, the aforementioned techniques should be considered analytical tools foremost, as the individual's privacy loss is itself a private quantity and releasing it requires special considerations [96].

6.2. **Privacy-Utility Trade-Offs.** As briefly discussed in section 4, DP in general adversely affects the utility of the model or of the results derived from a differentially private graph analytics. Utility if often measured by the accuracy of a query or with similar evaluation metrics. Therefore, similar to differentially private machine learning on Euclidean data or release of statistics derived from the sensitive data, there persists an issue of *privacy-utility trade-off*. This implies that the more "private" the result of the computation is (e.g. the lower the value of epsilon is), the less useful information can be inferred from that result not just by the adversary, but also by the end user of the trained model, potentially hindering the scientific progress based on the insights that could have otherwise been obtained from the study. This is further exacerbated by the inter-connected nature of the graphs, as it is not possible to guarantee independence of individual nodes, as we discussed above. Therefore, operations that limit the amount of information that can be derived from these nodes (e.g. through DP statistics release) affect not just the individuals, but also additional nodes connected to them. Thus, the utility loss can become more problematic when compared to datasets with independent data points and inflict additional penalties on the results of the computation. We note that this discussion is relevant to both graph datasets and GNNs, as the nature of GNN learning can only make full use of the data if these properties of graphs are preserved. Relying on GNN models pre-trained on publicly available data (similar to [42, 45, 93]) could severely reduce the negative impact that DP has on utility, when used in transfer learning contexts. Here a model is trained on public data and subsequently fine-tuned on private data where higher privacy can be achieved, while having better utility. This approach was demonstrated in [1] and more recently in [109] for non-graph machine learning tasks, demonstrating that –whereas training to the same utility *from scratch*– requires about one order of magnitude more data, results comparable to non-private training can easily be achieved by transfer learning.

6.3. **Computational Performance.** Beyond the aforementioned trade-offs in model generalisation performance, the utilisation of differential privacy is also associated with a computational performance overhead when employed in deep learning settings. This can be attributed to a requirement for per-sample gradient calculation, imposing a significant burden on model performance at train time. Moreover, due to noise addition and gradient clipping, models typically converge more slowly, thus prolonging the required training time [62].

Recent works have introduced new methods to reduce the shortcomings regarding computational performance of the DP training of ML models. Lee et al. [64] introduce a faster PyTorch implementation for per-sample gradient clipping for a variety of NN layers, including fully connected layers, recurrent layers, convolutional layers, LSTM layers, and multi-head attention layers. An extension to NLP methods called *ghost clipping* has been introduced by Li et al. [66]. Bu et al. [11] introduce a new book-keeping technique (alternative to ghost clipping) that sustainably improves computational costs of DP training and apply it to different convolutional networks. Also, deep learning frameworks supporting automatic vectorisation of per-sample gradient computations and/or just-in-time compilation can be used to greatly speed up the gradient clipping process [106]. We anticipate that a subset of these techniques can be straightforwardly extended to GNN training.

6.4. Interpretability of DP in Graphs.

DP can often be difficult to reason over from the perspectives of fairness [30] and explainability [26]. Moreover, its correct application is complicated by the introduction of unintuitive parameters like ε or δ [19, 53], or by the requirement to understand additional DP definitions like node, edge or graph-level DP. Thus, besides systems which automate sensitivity calculations and the application of DP to generic machine learning workflows [113], works similar to [18] are required, which investigate user expectations and interpretations of DP, paving the way for an improved *user experience* for practitioners.

Interpretability of GNNs in general is a highly discussed task in literature. The authors in [81] use an explainability method called *GNNExplainer* [125] to visualise and quantify the similarity between graph neural networks trained with and without DP-SGD to evaluate whether the privately trained network considers the same edges in the graph as important as the network trained with standard ML. We see potential in methods like these to get a better insight into differenitially private GNNs and increase their interpretability.

7. Conclusion

In this work, we explore and systematise the utilisation of differential privacy in methods that analyse graph data (including statistical analysis and graph neural network training) as well as graph generation methods. We discovered 57 works that perform differentially private data processing of graph structures, which we classify by the DP formulations employed in each work and summarise our findings in Tables 1 and 2. We identify three main DP formulations with regards to the attributes of graphs considered to be sensitive: (1) edge-level, (2) node-level, and (3) graph-level differential privacy. We additionally discuss machine learning tasks (in particular those relying on GNNs) that require utilisation of sensitive graph-structured data and could hence benefit from a formalisation of differentially private learning. Subsequently, we discuss the limitations of DP when applied to such learning contexts, some of which are inherent to the choice of DP learning setting and some attributable to the inter-connected nature of graph structures specifically. We conclude our discussion with an analysis of graph learning tasks on sensitive data, summarise which DP formulations are suitable for different learning problems and identify promising areas of future research. We hope that our work offers practitioners a helpful overview of the current state of DP employed in graph-based learning, and will stimulate both foundational and application-focused future research.

8. Acknowledgements

D.U. received funding from the TUM-Imperial College London Joint Academy of Doctoral Studies. R.B. received funding from the German Research Foundation (DFG) Special Priority Programme 2177 "Radiomics: next generation of biomedical imaging", project reference 428215948. D.R. and G.K. received funding from the UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare. The funders played no role in the design of the study, the preparation of the manuscript or the decision to publish.

References

[1] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016, pages 308–318. https://doi.org/10.1145/2976749.2978318.

[2] P. S. U. Adam Smith. Differentially private analysis on graphs, 2016. URL: https://cyber.biu.ac.il/wp-content/uploads/2016/09/graphs.pdf.

[3] R. Arora and J. Upadhyay. On differentially private graph sparsification and applications. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019, 32. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/e44e875c12109e4fa3716c05008048b2-Paper.pdf.

[4] B. Balle, G. Barthe, and M. Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 6280–6290, 2018, pages 6280–6290. URL: https://proceedings.neurips.cc/paper/2018/hash/3b5020bb891119b9f5130f1fea9bd773-Abstract.html.

[5] B. Balle, G. Barthe, M. Gaboardi, and J. Geumlek. Privacy amplification by mixing and diffusion mechanisms. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 13277–13287, 2019, pages 13277–13287. URL: https://proceedings.neurips.cc/paper/2019/hash/c4c42505a03f2e969b4c0a97ee9b34e7-Abstract.html.

[6] P. Barbiero, R. Viñas Torné, and P. Lió. Graph representation forecasting of patient's medical conditions: Toward a digital twin. *Frontiers in Genetics*, 12, 2021. URL: https://www.frontiersin.org/articles/10.3389/fgene.2021.652907, https://doi.org/10.3389/fgene.2021.652907.

[7] A. Benamira, B. Devillers, E. Lesot, A. K. Ray, M. Saadi, and F. D. Malliaros. Semi-supervised learning and graph neural networks for fake news detection. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 568–569. IEEE, 2019, pages 568–569. https://doi.org/10.1145/3341161.3342958.

[8] J. Blocki, A. Blum, A. Datta, and O. Sheffet. Differentially private data analysis of social networks via restricted sensitivity. In *Proceedings of the 4th conference*

on Innovations in Theoretical Computer Science, pages 87–96, 2013, pages 87–96. https://doi.org/10.1145/2422436.2422449.

[9] C. Borgs, J. T. Chayes, and A. D. Smith. Private graphon estimation for sparse graphs. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 1369–1377, 2015, pages 1369–1377. URL: https://proceedings.neurips.cc/paper/2015/hash/7250eb93b3c18cc9daa29cf58af7a004-Abstract.html.

[10] C. Borgs, J. T. Chayes, A. D. Smith, and I. Zadik. Revealing network structure, confidentially: Improved rates for node-private graphon estimation. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 533–543. IEEE Computer Society, 2018, pages 533–543. https://doi.org/10.1109/FOCS.2018.00057.

[11] Z. Bu, Y. Wang, S. Zha, and G. Karypis. Differentially private optimization on large model at small cost. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 3192–3218. PMLR, 2023, 202:3192–3218. URL: https://proceedings.mlr.press/v202/bu23a.html.

[12] E. T. Bullmore and D. S. Bassett. Brain graphs: Graphical models of the human brain connectome. *Annual Review of Clinical Psychology*, 7(1):113–140, 2011. PMID: 21128784. arXiv:https://doi.org/10.1146/annurev-clinpsy-040510-143934, https://doi.org/10.1146/annurev-clinpsy-040510-143934.

[13] R. Chen, B. C. Fung, S. Y. Philip, and B. C. Desai. Correlated network data publication via differential privacy. *The VLDB Journal*, 23(4):653–676, 2014. https://doi.org/10.1007/s00778-013-0344-8.

[14] S. Chen and S. Zhou. Recursive mechanism: towards node differential privacy and unrestricted joins. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, pages 653–664, 2013, pages 653–664. https://doi.org/10.1145/2463676.2465304.

[15] X. Chen, S. Mauw, and Y. Ramírez-Cruz. Publishing community-preserving attributed social graphs with a differential privacy guarantee. *Proc. Priv. Enhancing Technol.*, 2020(4):131–152, 2020. https://doi.org/10.2478/POPETS-2020-0066.

[16] A. Cheu, A. D. Smith, J. R. Ullman, D. Zeber, and M. Zhilyaev. Distributed differential privacy via shuffling. In Y. Ishai and V. Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part I*, volume 11476 of *Lecture Notes in Computer Science*, pages 375–403. Springer, 2019, 11476:375–403. https://doi.org/10.1007/978-3-030-17653-2_13.

[17] G. Cormode, S. Jha, T. Kulkarni, N. Li, D. Srivastava, and T. Wang. Privacy at scale: Local differential privacy in practice. In *Proceedings of the 2018 International Conference on Management of Data*, pages 1655–1658, 2018, pages 1655–1658.

[18] R. Cummings, V. Gupta, D. Kimpara, and J. Morgenstern. On the compatibility of privacy and fairness. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, UMAP'19 Adjunct, page 309–315, New York, NY, USA, 2019. Association for Computing Machinery, page 309–315. https://doi.org/10.1145/3314183.3323847.

[19] R. Cummings, G. Kaptchuk, and E. M. Redmiles. ”i need a better description”: An investigation into user expectations for differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, CCS ’21, page 3037–3052, New York, NY, USA, 2021. Association for Computing Machinery, page 3037–3052. https://doi.org/10.1145/3460120.3485252.

[20] A. Daigavane, G. Madan, A. Sinha, A. G. Thakurta, G. Aggarwal, and P. Jain. Node-level differentially private graph neural networks. *CoRR*, abs/2111.15521, 2021. URL: https://arxiv.org/abs/2111.15521, arXiv:2111.15521.

[21] W.-Y. Day, N. Li, and M. Lyu. Publishing graph degree distribution with node differential privacy. In *Proceedings of the 2016 International Conference on Management of Data*, pages 123–138, 2016, pages 123–138. https://doi.org/10.1145/2882903.2926745.

[22] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3(1):1–5, 2013. https://doi.org/10.1038/srep01376.

[23] D. Desfontaines and B. Pejó. Sok: Differential privacies. *Proc. Priv. Enhancing Technol.*, 2020(2):288–313, 2020. https://doi.org/10.2478/POPETS-2020-0028.

[24] J. Dong, A. Roth, and W. J. Su. Gaussian Differential Privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):3–37, 02 2022. arXiv:https://academic.oup.com/jrsssb/article-pdf/84/1/3/49324238/jrsssb_84_1_3.pdf, https://doi.org/10.1111/rssb.12454.

[25] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015, 28. URL: https://proceedings.neurips.cc/paper_files/paper/2015/file/f9be311e65d81a9ad8150a60844bb94c-Paper.pdf.

[26] C. Dwork, N. Kohli, and D. Mulligan. Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2), 2019. URL: https://par.nsf.gov/biblio/10217360, https://doi.org/10.29012/jpc.689.

[27] C. Dwork, A. Roth, et al. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014. https://doi.org/10.1561/0400000042.

[28] Ú. Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, K. Talwar, and A. Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2468–2479. SIAM, 2019, pages 2468–2479. https://doi.org/10.1137/1.9781611975482.151.

[29] H. Esfandiari, V. S. Mirrokni, U. Syed, and S. Vassilvitskii. Label differential privacy via clustering. *CoRR*, abs/2110.02159, 2021. URL: https://arxiv.org/abs/2110.02159, arXiv:2110.02159.

[30] T. Farrand, F. Mireshghallah, S. Singh, and A. Trask. Neither private nor fair: Impact of data imbalance on utility and fairness in differential privacy. In *Proceedings of the 2020 Workshop on Privacy-Preserving Machine Learning in Practice*, PPMLP’20, page 15–19, New York, NY, USA, 2020. Association for Computing Machinery, page 15–19. https://doi.org/10.1145/3411501.3419419.

[31] V. Feldman and T. Zrnic. Individual privacy accounting via a rényi filter. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 28080–28091. Curran Associates, Inc., 2021, 34:28080–28091. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/ec7f346604f518906d35ef0492709f78-Paper.pdf.

[32] H. Fichtenberger, M. Henzinger, and W. Ost. Differentially private algorithms for graphs under continual observation. *arXiv preprint arXiv:2106.14756*, 2021. https://doi.org/10.4230/LIPIcs.ESA.2021.42.

[33] U. Fiege, A. Fiat, and A. Shamir. Zero knowledge proofs of identity. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, page 210–217, New York, NY, USA, 1987. Association for Computing Machinery, page 210–217. https://doi.org/10.1145/28395.28419.

[34] M. Fredrikson, S. Jha, and T. Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, pages 1322–1333, 2015, pages 1322–1333. https://doi.org/10.1145/2810103.2813677.

[35] J. Gao, T. Lyu, F. Xiong, J. Wang, W. Ke, and Z. Li. Mgnn: A multimodal graph neural network for predicting the survival of cancer patients. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, page 1697–1700, New York, NY, USA, 2020. Association for Computing Machinery, page 1697–1700. https://doi.org/10.1145/3397271.3401214.

[36] T. Gaudelet, B. Day, A. R. Jamasb, J. Soman, C. Regep, G. Liu, J. B. Hayter, R. Vickers, C. Roberts, J. Tang, et al. Utilizing graph machine learning within drug discovery and development. *Briefings in bioinformatics*, 22(6):bbab159, 2021.

[37] J. Gehrke, E. Lui, and R. Pass. Towards privacy for social networks: A zero-knowledge based definition of privacy. In *Theory of cryptography conference*, pages 432–449. Springer, 2011, pages 432–449. https://doi.org/10.1007/978-3-642-19571-6_26.

[38] B. Ghazi, N. Golowich, R. Kumar, P. Manurangsi, and C. Zhang. On deep learning with label differential privacy. *CoRR*, abs/2102.06062, 2021. URL: https://arxiv.org/abs/2102.06062, arXiv:2102.06062.

[39] A. Ghosh and A. Roth. Selling privacy at auction. In *Proceedings 12th ACM Conference on Electronic Commerce (EC-2011), San Jose, CA, USA, June 5-9, 2011*, pages 199–208. ACM, 2011, pages 199–208. https://doi.org/10.1145/1993574.1993605.

[40] A. Gupta, A. Roth, and J. Ullman. Iterative constructions and private data release. In *Theory of cryptography conference*, pages 339–356. Springer, 2012, pages 339–356. https://doi.org/10.1007/978-3-642-28914-9_19.

[41] P. Hanzlicek, J. Spidlen, and M. Nagy. Universal electronic health record mudr. *Studies in health technology and informatics*, 105:190—201, 2004. URL: http://europepmc.org/abstract/MED/15718608.

[42] B. Hao, J. Zhang, H. Yin, C. Li, and H. Chen. Pre-training graph neural networks for cold-start users and items representation. In *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, pages 265–273. ACM, 2021, pages 265–273. https://doi.org/10.1145/3437963.3441738.

[43] M. Hay, C. Li, G. Miklau, and D. Jensen. Accurate estimation of the degree distribution of private networks. In *2009 Ninth IEEE International Conference on Data Mining*,

pages 169–178. IEEE, 2009, pages 169–178. https://doi.org/10.1109/ICDM.2009.11.

[44] X. He, R. Wen, Y. Wu, M. Backes, Y. Shen, and Y. Zhang. Node-level membership inference attacks against graph neural networks. *arXiv preprint arXiv:2102.05429*, 2021. https://doi.org/10.48550/arXiv.2102.05429.

[45] Z. Hu, Y. Dong, K. Wang, K. Chang, and Y. Sun. GPT-GNN: generative pre-training of graph neural networks. In R. Gupta, Y. Liu, J. Tang, and B. A. Prakash, editors, *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 1857–1867. ACM, 2020, pages 1857–1867. https://doi.org/10.1145/3394486.3403237.

[46] M. Iftikhar and Q. Wang. dk-projection: Publishing graph joint degree distribution with node differential privacy. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 358–370. Springer, 2021, pages 358–370. https://doi.org/10.1007/978-3-030-75765-6_29.

[47] T. Igamberdiev and I. Habernal. Privacy-preserving graph convolutional networks for text classification. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference, LREC 2022, Marseille, France, 20-25 June 2022*, pages 338–350. European Language Resources Association, 2022, pages 338–350. URL: https://aclanthology.org/2022.lrec-1.36.

[48] J. Imola, T. Murakami, and K. Chaudhuri. Locally differentially private analysis of graph statistics. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 983–1000. USENIX Association, 2021, pages 983–1000. URL: https://www.usenix.org/conference/usenixsecurity21/presentation/imola.

[49] D. Jiang, Z. Wu, C. Hsieh, G. Chen, B. Liao, Z. Wang, C. Shen, D. Cao, J. Wu, and T. Hou. Could graph neural networks learn better molecular representation for drug discovery? a comparison study of descriptor-based and graph-based models. *J. Cheminformatics*, 13(1):12, 2021. https://doi.org/10.1186/S13321-020-00479-8.

[50] J. Jordon, J. Yoon, and M. van der Schaar. PATE-GAN: generating synthetic data with differential privacy guarantees. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL: https://openreview.net/forum?id=S1zk9iRqF7.

[51] Z. Jorgensen, T. Yu, and G. Cormode. Publishing attributed social graphs with formal privacy guarantees. In *Proceedings of the 2016 international conference on management of data*, pages 107–122, 2016, pages 107–122. https://doi.org/10.1145/2882903.2915215.

[52] P. Kairouz, K. Bonawitz, and D. Ramage. Discrete distribution estimation under local privacy. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2436–2444, New York, New York, USA, 20–22 Jun 2016. PMLR, 48:2436–2444. URL: https://proceedings.mlr.press/v48/kairouz16.html.

[53] G. Kaissis, M. Knolle, F. Jungmann, A. Ziller, D. Usynin, and D. Rueckert. A unified interpretation of the gaussian mechanism for differential privacy through the sensitivity index. *Journal of Privacy and Confidentiality*, 12(1), Jul. 2022. URL: https://journalprivacyconfidentiality.org/index.php/jpc/article/view/807, https://doi.org/10.29012/jpc.807.

[54] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine*

Intelligence, 2(6):305–311, 2020. https://doi.org/10.1038/s42256-020-0186-1.

[55] V. Karwa, S. Raskhodnikova, A. Smith, and G. Yaroslavtsev. Private analysis of graph structure. *Proceedings of the VLDB Endowment*, 4(11):1146–1157, 2011. https://doi.org/10.14778/3402707.3402749.

[56] V. Karwa and A. B. Slavković. Differentially private graphical degree sequences and synthetic graphs. In *International Conference on Privacy in Statistical Databases*, pages 273–285. Springer, 2012, pages 273–285. https://doi.org/10.1007/978-3-642-33627-0_21.

[57] V. Karwa and A. Slavković. Inference using noisy degrees: Differentially private β-model and synthetic graphs. *The Annals of Statistics*, 44(1):87 – 112, 2016. https://doi.org/10.1214/15-AOS1358.

[58] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011. arXiv:https://doi.org/10.1137/090756090, https://doi.org/10.1137/090756090.

[59] S. P. Kasiviswanathan, K. Nissim, S. Raskhodnikova, and A. Smith. Analyzing graphs with node differential privacy. In *Theory of Cryptography Conference*, pages 457–476. Springer, 2013, pages 457–476. https://doi.org/10.1007/978-3-642-36594-2_26.

[60] M. Kessel, P. Ruppel, and F. Gschwandtner. BIGML: A location model with individual waypoint graphs for indoor location-based services. *Prax. Inf.verarb. Kommun.*, 33(4):261–267, 2010. https://doi.org/10.1515/PIKO.2010.045.

[61] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL: https://openreview.net/forum?id=SJU4ayYgl.

[62] A. Kurakin, S. Chien, S. Song, R. Geambasu, A. Terzis, and A. Thakurta. Toward training at imagenet scale with differential privacy. *CoRR*, abs/2201.12328, 2022. URL: https://arxiv.org/abs/2201.12328, arXiv:2201.12328.

[63] S. Lan, H. Xin, W. Yingjie, and G. Yongyi. Sensitivity reduction of degree histogram publication under node differential privacy via mean filtering. *Concurrency and Computation: Practice and Experience*, 33(8):e5621, 2021. https://doi.org/10.1002/cpe.5621.

[64] J. Lee and D. Kifer. Scaling up differentially private deep learning with fast per-example gradient clipping. *Proc. Priv. Enhancing Technol.*, 2021(1):128–144, 2021. https://doi.org/10.2478/POPETS-2021-0008.

[65] X. Li, N. C. Dvornek, Y. Zhou, J. Zhuang, P. Ventola, and J. S. Duncan. Graph neural network for interpreting task-fmri biomarkers. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 485–493. Springer International Publishing, 2019, pages 485–493.

[66] X. Li, F. Tramèr, P. Liang, and T. Hashimoto. Large language models can be strong differentially private learners. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL: https://openreview.net/forum?id=bVuP3ltATMz.

[67] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola, and J. S. Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis*, 74:102233, 2021. URL: https://www.sciencedirect.com/science/article/pii/S1361841521002784, https://doi.org/https://doi.org/10.1016/j.media.2021.102233.

[68] C. Liu, S. Chakraborty, and P. Mittal. Dependence makes you vulnerable: differential privacy under dependent tuples. In *NDSS*, volume 16, pages 21–24, 2016, 16:21–24. https://doi.org/10.14722/ndss.2016.23279.

[69] C. Liu, F. Wang, J. Hu, and H. Xiong. Temporal phenotyping from longitudinal electronic health records: A graph based framework. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, page 705–714, New York, NY, USA, 2015. Association for Computing Machinery, page 705–714. https://doi.org/10.1145/2783258.2783352.

[70] G. Liu, X. Ma, and W. Li. Publishing node strength distribution with node differential privacy. *IEEE Access*, 8:217642–217650, 2020. https://doi.org/10.1109/ACCESS.2020.3040077.

[71] W. Liu, B. Liu, Q. Xu, and H. Lei. Graph node strength histogram publication method with node differential privacy. In *Journal of Physics: Conference Series*, volume 1757, page 012186. IOP Publishing, 2021, 1757:012186. https://doi.org/10.1088/1742-6596/1757/1/012186.

[72] W. Lu and G. Miklau. Exponential random graph estimation under differential privacy. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 921–930, 2014, pages 921–930. https://doi.org/10.1145/2623330.2623683.

[73] A. Machanavajjhala, A. Korolova, and A. Sarma. Personalized social recommendations accurate or private? *Proceedings of the VLDB Endowment*, 4(7):440–450, Apr. 2011. https://doi.org/10.14778/1988776.1988780.

[74] K. R. Macwan and S. J. Patel. Node differential privacy in social graph degree publishing. *Procedia computer science*, 143:786–793, 2018. https://doi.org/10.1016/j.procs.2018.10.388.

[75] M. Malek, I. Mironov, K. Prasad, I. Shilov, and F. Tramèr. Antipodes of label differential privacy: PATE and ALIBI. *CoRR*, abs/2106.03408, 2021. URL: https://arxiv.org/abs/2106.03408, arXiv:2106.03408.

[76] C. Mao, L. Yao, and Y. Luo. Medgcn: Graph convolutional networks for multiple medical tasks. *arXiv preprint arXiv:1904.00326*, 2019. https://doi.org/10.1016/j.jbi.2022.104000.

[77] D. Matsunaga, T. Suzumura, and T. Takahashi. Exploring graph neural networks for stock market predictions with rolling window analysis. Papers, arXiv.org, 2019. URL: https://EconPapers.repec.org/RePEc:arx:papers:1909.10660.

[78] D. Mir and R. N. Wright. A differentially private estimator for the stochastic kronecker graph model. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, pages 167–176, 2012, pages 167–176. https://doi.org/10.1145/2320765.2320818.

[79] D. J. Mir and R. N. Wright. A differentially private graph estimator. In *2009 IEEE International Conference on Data Mining Workshops*, pages 122–129. IEEE, 2009, pages 122–129. https://doi.org/10.1109/ICDMW.2009.96.

[80] I. Mironov. Rényi differential privacy. In *30th IEEE Computer Security Foundations Symposium, CSF 2017, Santa Barbara, CA, USA, August 21-25, 2017*, pages 263–275. IEEE Computer Society, 2017, pages 263–275. https://doi.org/10.1109/CSF.2017.11.

[81] T. T. Mueller, J. C. Paetzold, C. Prabhakar, D. Usynin, D. Rueckert, and G. Kaissis. Differentially private graph neural networks for whole-graph classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–11, 2022. https:

//doi.org/10.1109/TPAMI.2022.3228315.

[82] M. Ménoret, N. Farrugia, B. Pasdeloup, and V. Gripon. Evaluating graph signal processing for neuroimaging through classification and dimensionality reduction. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 618–622, 2017, pages 618–622. https://doi.org/10.1109/GlobalSIP.2017.8309033.

[83] R. Müller, O. Thews, C. Rohrbach, M. Sergl, and K. Pommerening. A graph-grammar approach to represent causal, temporal and other contexts in an oncological patient record. *Methods of information in medicine*, 35(2):127—141, June 1996. URL: http://europepmc.org/abstract/MED/8755386.

[84] K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007, pages 75–84. https://doi.org/10.1145/1250790.1250803.

[85] I. E. Olatunji, T. Funke, and M. Khosla. Releasing graph neural networks with differential privacy guarantees. *Transactions on Machine Learning Research*, 2023. URL: https://openreview.net/forum?id=wk8oXR0kFA.

[86] I. E. Olatunji, W. Nejdl, and M. Khosla. Membership inference attack on graph neural networks. In *2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, pages 11–20, 2021, pages 11–20. https://doi.org/10.1109/TPSISA52974.2021.00002.

[87] N. Papernot, M. Abadi, Ú. Erlingsson, I. Goodfellow, and K. Talwar. Semi-supervised knowledge transfer for deep learning from private training data. In *International Conference on Learning Representations*, 2017. URL: https://openreview.net/forum?id=HkwoSDPgg.

[88] N. Papernot, S. Song, I. Mironov, A. Raghunathan, K. Talwar, and Ú. Erlingsson. Scalable private learning with PATE. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL: https://openreview.net/forum?id=rkZB1XbRZ.

[89] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert. Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer's disease. *Medical Image Analysis*, 48:117–130, 2018. URL: https://www.sciencedirect.com/science/article/pii/S1361841518303554, https://doi.org/https://doi.org/10.1016/j.media.2018.06.001.

[90] W. Paul, Y. Cao, M. Zhang, and P. Burlina. Defending medical image diagnostics against privacy attacks using generative methods: Application to retinal diagnostics. In *Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning - 10th Workshop, CLIP 2021, Second Workshop, DCL 2021, First Workshop, LL-COVID19 2021, and First Workshop and Tutorial, PPML 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27 and October 1, 2021, Proceedings*, volume 12969 of *Lecture Notes in Computer Science*, pages 174–187. Springer, 2021, 12969:174–187. https://doi.org/10.1007/978-3-030-90874-4_17.

[91] D. Proserpio, S. Goldberg, and F. McSherry. Calibrating data to sensitivity in private data analysis: A platform for differentially-private analysis of weighted datasets.

Proc. VLDB Endow., 7(8):637–648, apr 2014. https://doi.org/10.14778/2732296.2732300.

[92] Z. Qin, T. Yu, Y. Yang, I. Khalil, X. Xiao, and K. Ren. Generating synthetic decentralized social graphs with local differential privacy. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 425–438, 2017, pages 425–438. https://doi.org/10.1145/3133956.3134086.

[93] J. Qiu, Q. Chen, Y. Dong, J. Zhang, H. Yang, M. Ding, K. Wang, and J. Tang. GCC: graph contrastive coding for graph neural network pre-training. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 1150–1160. ACM, 2020, pages 1150–1160. https://doi.org/10.1145/3394486.3403168.

[94] S. Raskhodnikova and A. Smith. *Private Analysis of Graph Data*, pages 1–6. Springer Berlin Heidelberg, Berlin, Heidelberg. pages 1–6, 2014. https://doi.org/10.1007/978-3-642-27848-8_549-1.

[95] S. Raskhodnikova and A. Smith. Lipschitz extensions for node-private graph statistics and the generalized exponential mechanism. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 495–504. IEEE, 2016, pages 495–504. https://doi.org/10.1109/FOCS.2016.60.

[96] R. Redberg and Y. Wang. Privately publishable per-instance privacy. In M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang, and J. W. Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 17335–17346, 2021, pages 17335–17346. URL: https://proceedings.neurips.cc/paper/2021/hash/9087b0efc7c7acd1ef7e153678809c77-Abstract.html.

[97] J. Richiardi, S. Achard, H. Bunke, and D. Van De Ville. Machine learning with brain graphs: predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal processing magazine*, 30(3):58–70, 2013. https://doi.org/10.1109/MSP.2012.2233865.

[98] S. Sajadmanesh and D. Gatica-Perez. Locally private graph neural networks. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, CCS '21, page 2130–2145, New York, NY, USA, 2021. Association for Computing Machinery, page 2130–2145. https://doi.org/10.1145/3460120.3484565.

[99] S. Sajadmanesh, A. S. Shamsabadi, A. Bellet, and D. Gatica-Perez. GAP: differentially private graph neural networks with aggregation perturbation. In J. A. Calandrino and C. Troncoso, editors, *32nd USENIX Security Symposium, USENIX Security 2023, Anaheim, CA, USA, August 9-11, 2023*, pages 3223–3240. USENIX Association, 2023, pages 3223–3240. URL: https://www.usenix.org/conference/usenixsecurity23/presentation/sajadmanesh.

[100] A. Sala, X. Zhao, C. Wilson, H. Zheng, and B. Y. Zhao. Sharing graphs using differentially private graph models. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, page 81–98, New York, NY, USA, 2011. Association for Computing Machinery, page 81–98. https://doi.org/10.1145/2068816.2068825.

[101] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009. https://doi.org/10.1109/TNN.2008.2005605.

[102] A. Sealfon. Shortest paths and distances with differential privacy. In *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 29–41. ACM, 2016, pages 29–41. https://doi.org/10.1145/2902251.2902291.

[103] E. Shen and T. Yu. Mining frequent graph patterns with differential privacy. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 545–553, 2013, pages 545–553. https://doi.org/10.1145/2487575.2487601.

[104] R. Shokri, M. Stronati, C. Song, and V. Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3–18. IEEE, 2017, pages 3–18. https://doi.org/10.1109/SP.2017.41.

[105] O. Sporns. From simple graphs to the connectome: Networks in neuroimaging. *NeuroImage*, 62(2):881–886, 2012. 20 YEARS OF fMRI. URL: https://www.sciencedirect.com/science/article/pii/S1053811911010172, https://doi.org/https://doi.org/10.1016/j.neuroimage.2011.08.085.

[106] P. Subramani, N. Vadivelu, and G. Kamath. Enabling fast differentially private SGD via just-in-time compilation and vectorization. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 26409–26421, 2021, pages 26409–26421. URL: https://proceedings.neurips.cc/paper/2021/hash/ddf9029977a61241841edeae15e9b53f-Abstract.html.

[107] H. Sun, X. Xiao, I. Khalil, Y. Yang, Z. Qin, W. H. Wang, and T. Yu. Analyzing subgraph statistics from extended local views with decentralized differential privacy. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019, London, UK, November 11-15, 2019*, pages 703–717. ACM, 2019, pages 703–717. https://doi.org/10.1145/3319535.3354253.

[108] C. Task and C. Clifton. What should we protect? defining differential privacy for social network analysis. In *State of the Art Applications of Social Network Analysis*, pages 139–161. Springer, 2014. pages 139–161. https://doi.org/10.1007/978-3-319-05912-9_7.

[109] F. Tramèr and D. Boneh. Differentially private learning needs better features (or much more data). In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL: https://openreview.net/forum?id=YTWGvpFOQD-.

[110] M. C. Tschantz, S. Sen, and A. Datta. Sok: Differential privacy as a causal property. In *2020 IEEE Symposium on Security and Privacy, SP 2020, San Francisco, CA, USA, May 18-21, 2020*, pages 354–371. IEEE, 2020, pages 354–371. https://doi.org/10.1109/SP40000.2020.00012.

[111] J. Ullman and A. Sealfon. Efficiently estimating erdos-renyi graphs with node differential privacy. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019, 32. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/955cb567b6e38f4c6b3f28cc857fc38c-Paper.pdf.

[112] J. R. Ullman and A. Sealfon. Efficiently estimating erdos-renyi graphs with node differential privacy. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 3765–3775,

2019, pages 3765–3775. URL: https://proceedings.neurips.cc/paper/2019/hash/955cb567b6e38f4c6b3f28cc857fc38c-Abstract.html.

[113] D. Usynin, A. Ziller, M. Knolle, D. Rueckert, and G. Kaissis. An automatic differentiation system for the age of differential privacy. *CoRR*, abs/2109.10573, 2021. URL: https://arxiv.org/abs/2109.10573, arXiv:2109.10573.

[114] Y. Wang, B. Balle, and S. P. Kasiviswanathan. Subsampled rényi differential privacy and analytical moments accountant. *J. Priv. Confidentiality*, 10(2), 2020. https://doi.org/10.29012/JPC.723.

[115] Y. Wang and X. Wu. Preserving differential privacy in degree-correlation based graph generation. *Trans. Data Priv.*, 6(2):127–145, 2013. URL: http://www.tdp.cat/issues11/abs.a113a12.php.

[116] Y. Wang, X. Wu, and D. Hu. Using randomized response for differential privacy preserving data collection. In *EDBT/ICDT Workshops*, 2016. URL: https://api.semanticscholar.org/CorpusID:1286991.

[117] Y. Wang, X. Wu, and L. Wu. Differential privacy preserving spectral graph analysis. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 329–340. Springer, 2013, pages 329–340. https://doi.org/10.1007/978-3-642-37456-2_28.

[118] Y.-X. Wang. Per-instance differential privacy. *Journal of Privacy and Confidentiality*, 9(1), Mar. 2019. URL: https://journalprivacyconfidentiality.org/index.php/jpc/article/view/662, https://doi.org/10.29012/jpc.662.

[119] X. Wei, R. Yu, and J. Sun. View-gcn: View-based graph convolutional network for 3d shape analysis. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1847–1856, 2020, pages 1847–1856. https://doi.org/10.1109/CVPR42600.2020.00192.

[120] F. Wu, Y. Long, C. Zhang, and B. Li. LINKTELLER: recovering private edges from graph neural networks via influence analysis. In *43rd IEEE Symposium on Security and Privacy, SP 2022, San Francisco, CA, USA, May 22-26, 2022*, pages 2005–2024. IEEE, 2022, pages 2005–2024. https://doi.org/10.1109/SP46214.2022.9833806.

[121] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2021. https://doi.org/10.1109/TNNLS.2020.2978386.

[122] S. Xia, B. Chang, K. Knopf, Y. He, Y. Tao, and X. He. Dpgraph: A benchmark platform for differentially private graph analysis. In *SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021*, pages 2808–2812. ACM, 2021, pages 2808–2812. https://doi.org/10.1145/3448016.3452756.

[123] M. Xie, H. Yin, H. Wang, F. Xu, W. Chen, and S. Wang. Learning graph-based poi embedding for location-based recommendation. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, page 15–24, New York, NY, USA, 2016. Association for Computing Machinery, page 15–24. https://doi.org/10.1145/2983323.2983711.

[124] X. Xiong, S. Liu, D. Li, Z. Cai, and X. Niu. A comprehensive survey on local differential privacy. *Security and Communication Networks*, 2020, 2020. https://doi.org/10.1155/2020/8829523.

[125] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec. Gnn explainer: A tool for post-hoc explanation of graph neural networks. *CoRR*, abs/1903.03894, 2019. URL: http://arxiv.org/abs/1903.03894, arXiv:1903.03894.

[126] S. Yovine, F. Mayr, S. Sosa, and R. Visca. An assessment of the application of private aggregation of ensemble models to sensible data. *Machine Learning and Knowledge Extraction*, 3(4):788–801, 2021. URL: https://www.mdpi.com/2504-4990/3/4/39, https://doi.org/10.3390/make3040039.

[127] Q. Yu, E. B. Erhardt, J. Sui, Y. Du, H. He, D. Hjelm, M. S. Cetin, S. Rachakonda, R. L. Miller, G. Pearlson, and V. D. Calhoun. Assessing dynamic brain graphs of time-varying connectivity in fmri data: Application to healthy controls and patients with schizophrenia. *NeuroImage*, 107:345–355, 2015. URL: https://www.sciencedirect.com/science/article/pii/S105381191401012X, https://doi.org/https://doi.org/10.1016/j.neuroimage.2014.12.020.

[128] Z. Yuxuan, W. Jianghong, L. Ji, L. Wenfen, and H. Xuexian. Graph degree histogram publication method with node-differential privacy. *Journal of Computer Research and Development*, 56(3):508, 2019. https://doi.org/10.1145/2882903.2926745.

[129] H. Zhang, S. Latif, R. Bassily, and A. Rountev. Differentially-private control-flow node coverage for software usage analysis. In *Proceedings of the 29th USENIX Conference on Security Symposium*, SEC'20, USA, 2020. USENIX Association.

[130] Q. Zhang, J. Ma, J. Lou, L. Xiong, and X. Jiang. Towards training robust private aggregation of teacher ensembles under noisy labels. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 1103–1110, 2020, pages 1103–1110. https://doi.org/10.1109/BigData50022.2020.9378234.

[131] S. Zhang, W. Ni, and N. Fu. Community preserved social graph publishing with node differential privacy. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 1400–1405. IEEE, 2020, pages 1400–1405. https://doi.org/10.1109/ICDM50108.2020.00184.

[132] Z. Zhang, Q. Liu, Z. Huang, H. Wang, C. Lu, C. Liu, and E. Chen. Graphmi: Extracting private graph data from graph neural networks. In Z.-H. Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 3749–3755. International Joint Conferences on Artificial Intelligence Organization, 8 2021, pages 3749–3755. Main Track. https://doi.org/10.24963/ijcai.2021/516.

[133] X. Zheng, N. McCarthy, and J. Hayes. Network generation with differential privacy. *arXiv preprint arXiv:2111.09085*, 2021. https://doi.org/10.48550/arXiv.2111.09085.

[134] X. Zheng, L. Zhang, K. Li, and X. Zeng. Efficient publication of distributed and overlapping graph data under differential privacy. *Tsinghua Science and Technology*, 27(2):235–243, 2021. https://doi.org/10.26599/TST.2021.9010018.

[135] H. Zhu, X. Zuo, and M. Xie. DP-FT: A differential privacy graph generation with field theory for social network data release. *IEEE Access*, 7:164304–164319, 2019. https://doi.org/10.1109/ACCESS.2019.2952452.

[136] T. Zhu, G. Li, W. Zhou, and S. Y. Philip. *Differential privacy and applications*. Springer, 2017. https://doi.org/10.1007/978-3-319-62004-6.

This work is licensed under the Creative Commons License Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0/ or send a letter to Creative Commons, 171 Second St, Suite 300, San Francisco, CA 94105, USA, or Eisenacher Strasse 2, 10777 Berlin, Germany