Continuous Publication of Weighted Graphs with Local Differential Privacy

Wen Xu
Jinan University
wenxu018@gmail.com

Pengpeng Qiao
Institute of Science Tokyo
peng2qiao@gmail.com

Shang Liu
CUMT
shang@cumt.edu.cn

Zhirun Zheng
Ajou University
zhengzhirun2019@gmail.com

Yang Cao
Institute of Science Tokyo
cao@c.titech.ac.jp

Zhetao Li
Jinan University
liztchina@gmail.com

ABSTRACT

Although a large amount of valuable knowledge can be obtained from the weighted graph snapshots modeled over time, it may cause privacy issues. Local differential privacy (LDP) provides a strong solution for private graph data publishing in decentralized networks. However, most existing LDP studies over graphs are only applicable to static unweighted graphs. This paper investigates the problem of continuous publication of weighted graph snapshots and proposes a graph publication framework, WGT-LDP, under w-event edge weight LDP, which can protect the privacy of edges and weights over any w consecutive time steps. WGT-LDP consists of four key components: population division-based sampling that overcomes the problem of over-segmentation of the privacy budget, data range estimation that mitigates noise on edge weights, aggregate information collection that obtains important information about the graph structure and edge weights, and graph snapshot generation that reconstructs weighted graph snapshot at each time step. We provide theoretical guarantees on privacy and utility, and perform extensive experiments on three real-world and two synthetic datasets, using four commonly used metrics. Our experiments show that WGT-LDP produces high-quality synthetic weighted graphs and significantly outperforms baseline methods.

PVLDB Reference Format:
Wen Xu, Pengpeng Qiao, Shang Liu, Zhirun Zheng, Yang Cao, and Zhetao Li. Continuous Publication of Weighted Graphs with Local Differential Privacy. PVLDB, 14(1): XXX-XXX, 2020.
doi:XX.XX/XXX.XX

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/xuwen22/WGT-LDP.

1 INTRODUCTION

Due to graphs providing an excellent ability to represent relational data, they have been widely used in many real-world complex systems, such as social media and computer networks [26]. In many cases, the interaction behaviors between entities in these systems often change dynamically over time, which can be modeled as continuous graph snapshots. Moreover, these snapshots may carry more valuable information beyond connecting different entities (called weights), e.g., interaction frequency, which can be collected and analyzed for further complex tasks. The following are two real examples of publishing and analyzing weighted graph snapshots.

Example 1. ***Social Network Analytics***. *On professional social platforms such as LinkedIn, the edges and weights in weighted graph snapshots can reflect the actual dynamic relevance and connection strength between users, thereby increasing the efficiency of job search and recruitment [15].*

Example 2. ***Epidemiological Network Analytics***. *Analyzing continuous weighted graph data is also great popularity in epidemiological network systems, e.g., respiratory infectious diseases. By studying the dynamic interactions (edges) and interaction strengths (weights) between individuals in continuous snapshots, their risk of contracting respiratory diseases can be estimated [10, 12].*

However, directly publishing these graph data for analysis without protection may pose privacy risks. Local differential privacy (LDP) [7, 22] is a strong privacy-preserving technique that can be used to collect sensitive data from users in a decentralized network without a complete central database. In the LDP setting, to ensure privacy, each user must send the data injected with random noise to the curator. Existing work on LDP-based graph publishing mainly focuses on static unweighted graphs [35, 44], while very little focus on weighted graph data in the temporal dimension. Despite this, providing privacy for dynamic networks with abundant valuable information may be more common in the near future [40].

More precisely, our scenario is concerned with the continuous publication of weighted graph snapshots, where both the graph structure and edge weights may be private information. We focus on methods that provide w-event-level privacy because they are applicable to infinite time steps while providing strong privacy protection [23]. In particular, w-event-level privacy aims to protect any event sequence occurring within any window of w time steps, which means that individuals can guarantee the privacy of their data over w consecutive snapshots. Figure 1 shows an example on a decentralized network where the curator regularly collects perturbation data from all users to continuously synthesize weighted graphs. To the best of our knowledge, previous works on w-event-level privacy have focused on tabular streaming data [28, 30, 36, 43], and have not explored continuous weighted graph publication under LDP. To this end, we have three technical challenges:

Proceedings of the VLDB Endowment, Vol. 14, No. 1 ISSN 2150-8097.
doi:XX.XX/XXX.XX

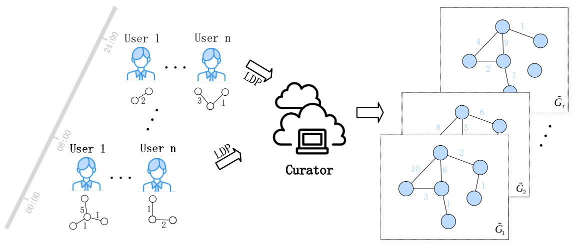


Figure 1: The example of continuously synthesizing weighted graph snapshots in decentralized networks.

- **Over-segmentation of privacy budget.** To achieve w-event-level privacy, most existing methods on tabular data need to allocate privacy budgets for any sliding window of size w, so that the sum of the budgets within this window does not exceed the total privacy budget ε. However, the publishing mechanism of graph data usually contains multiple components that consume privacy budgets, which will cause further segmentation of the privacy budget and thus destroy the utility of the data.
- **High noise magnitude of edge weights.** In order not to violate the differential privacy protocol, the magnitude of the noise added to the edge weight should be proportional to its upper bound. In the study of central DP [46], the upper bound can be lowered based on the total error of all weights in the original graph. This is difficult under LDP since the curator cannot access the original data.
- **Bias of noisy adjacency matrix.** After adding noise to every bit of the adjacency matrix with weights, zero edges are converted to edges with real weights, making the results no longer useful. It is challenging to mitigate the negative impact of noise on the performance of synthetic weighted graph snapshots.

In this paper, we propose WGT-LDP, a new framework for continuous weighted graph publishing under w-event edge weight LDP. We first design a population division-based sampling scheme, which considers data changes and samples disjoint nodes in the window, so that each node has sufficient privacy budget to perturb its data. To effectively reduce the noise of edge weights, we propose a data range estimation mechanism where the curator collects the noisy local maximum weights of sampling nodes and runs a post-processing technique to estimate the local data range of each node. Since node degrees and adjacency lists carry knowledge about graph structure and edge weights, respectively, we collect these two aggregates of information from sampling nodes. By using the coarse-grained node noisy degrees as evidence and the fine-grained noisy adjacency lists as reference, we can recover the characteristics of the original snapshot and preserve structural sparsity. Our main contributions are summarized as follows:

New Perspective. We first explore the problem of weighted graph publication with LDP in the temporal dimension, which aims to generate continuous synthetic weighted graph snapshots while satisfying privacy protection requirements.

Simple yet Effective Solution. We propose a continuous weighted graphs publication framework WGT-LDP under w-event edge weight LDP. WGT-LDP first adopts a population division strategy to adaptively sample nodes with large data changes, and then uses a data range estimation method to reduce the impact of perturbation noise on edge weights. Finally, WGT-LDP collects noisy degrees as evidence and noisy adjacency lists as reference and adopts different methods to reconstruct the current snapshot to ensure weighted graph utility.

Extensive Experimental Evaluations. We theoretically prove that the proposed WGT-LDP satisfies w-event edge weight LDP and analyze the utility. Extensive experiments on several datasets and metrics demonstrate the effectiveness of WGT-LDP.

2 PRELIMINARIES

2.1 Local Differential Privacy (LDP)

DP [8], a standard for privacy preserving data publication, originally assumed a centralized model where a trusted curator holds the exact data of all users. This may lead to some privacy and security issues. For instance, a data curator may sell data for personal gain or suffer an attack that leads to data leakage [38]. Local differential privacy (LDP) [7] effectively solves the above problems by not assuming a trusted third party. In LDP, each user can use the DP mechanism to perturb personal sensitive data locally before the data is sent to the curator. Formally, LDP is defined as follows:

Definition 2.1 (LDP). A randomized algorithm $\mathcal{M}$ provides ε-LDP, where $\varepsilon > 0$, if and only if for any pair of input values $x, x' \in D$ and any possible output x^*,

$$\Pr\left[\mathcal{M}(x) = x^*\right] \le e^{\varepsilon} \Pr\left[\mathcal{M}(x') = x^*\right]. \tag{1}$$

Intuitively, the probability that an attacker can distinguish whether the original data is x or x' by observing the output of $\mathcal{M}$ is controlled by ε, which is called the privacy budget. Therefore, the smaller ε can provide stronger privacy guarantees.

Geometric Mechanism. The Geometric mechanism (GM) [13] is mainly used for queries with integer results. It satisfies the LDP requirements by adding random geometric noise to the query function f on the input x. The magnitude of noise is proportional to the global sensitivity, defined as,

$$\Delta f = \underset{x,x' \in D}{\text{maximize}} \|f(x) - f(x')\|_1. \tag{2}$$

For numerical data, the output of GM $\mathcal{M}$ is as follows:

$$\mathcal{M}(x) = f(x) + Geo(e^{-\varepsilon/\Delta f}), \tag{3}$$

where $Geo(\lambda)$ denotes a random integer noise drawn from two-sided geometric distribution $\Pr[Geo(\lambda) = z] = \frac{1-\lambda}{1+\lambda}\lambda^{|z|}$, and it has a mean of 0.

Square Wave Mechanism. The Square Wave mechanism (SW) [27] attempts to increase the probability that the noisy response value carries useful information about the true value, which extends the idea of generalized randomized response. That is, the probability that users report values closer to x is greater than values farther away from x for given input x.

In particular, we assume that the input domain is $[0, 1]$ and the output domain is $[-b, 1 + b]$, where $b = \frac{\varepsilon e^{\varepsilon} - e^{\varepsilon} + 1}{2e^{\varepsilon}(e^{\varepsilon} - 1 - \varepsilon)}$. For values in the range $[l, h]$, they are first mapped into $[0, 1]$ (by transforming value x to $\frac{x-l}{h-l}$), and then the estimated values are transformed back.

Algorithm 1 EMS algorithm [27]

Input: Perturbed values $\tilde{\mathbf{x}}$, transformation probabilities M, bucket size r

Output: Probability distribution $\tilde{\mathbf{z}}$

Initialize $\tilde{\mathbf{z}} = \{1/r, 1/r, \cdots, 1/r\}$;

while not converge **do**

  $\forall i \in \{1, 2, \cdots, r\}, p_i = \tilde{\mathbf{z}}_i \sum_{j \in [r]} n_j \frac{M_{j,i}}{\sum_{k=1}^{r} M_{j,k} \tilde{\mathbf{z}}_k}$; // *E-step*

  $\forall i \in \{1, 2, \cdots, r\}, \tilde{\mathbf{z}}_i = \frac{p_i}{\sum_{k=1}^{r} p_k}$; // *M-step*

  $\tilde{\mathbf{z}}_i = \frac{1}{2} \tilde{\mathbf{z}}_i + \frac{1}{4} \left(\tilde{\mathbf{z}}_{i-1} + \tilde{\mathbf{z}}_{i+1}\right)$; // *S-step*

end while

return $\tilde{\mathbf{z}}$

The perturbation function of SW is defined as:

$$\forall_{\tilde{x} \in [-b, 1+b]}, \Pr[\mathrm{SW}(x) = \tilde{x}] = \begin{cases} p, & \text{if } |x - \tilde{x}| \leq b, \\ q, & \text{otherwise}, \end{cases} \tag{4}$$

where $p = \frac{e^{\varepsilon}}{2be^{\varepsilon}+1}$ and $q = \frac{1}{2be^{\varepsilon}+1}$.

The LDP mechanisms satisfy the properties of sequential/parallel composition and post-processing [8, 31], which provide privacy guarantees for building complex LDP algorithms.

Expectation Maximization with Smoothing Algorithm. The Expectation Maximization with Smoothing (EMS) algorithm [27] is a post-processing technique for estimating the original distribution of noisy reports, which is shown in Algorithm 1. In particular, the integer r represents the bucket size obtained after the numeric domain is discretized using binning. The matrix M represents the transformation probabilities, where $M_{j,k}$ denotes the probability that the output value falls into bucket $b_j, j \in [r]$, given the input in bucket $b_k, k \in [r]$. The input data is assumed to be uniformly distributed in bucket b_k. Each column of the matrix M sums up to 1 [1].

2.2 LDP for Weighted Graphs

In the context of weighted graphs, both the edges and edge weights may be private information. Depending on the privacy requirement, we introduce a formal definition of LDP for weighted graphs.

Specifically, let $V = \{v_1, v_2, \cdots, v_n\}$ be the set of all nodes (i.e., users) in a undirected weighted graph. The adjacency list of a node v_i can be denoted as an n-dimensional bit vector $(a(v_i, v_1), \cdots, a(v_i, v_n))$, where $a(v_i, v_j)$ is the edge weight between nodes v_i and v_j. Note that if edge (v_i, v_j) exists in the weighted graph, then $a(v_i, v_j) \geq 1$; otherwise $a(v_i, v_j) = 0$. The LDP for weighted graphs is formally defined as follows:

Definition 2.2 (edge weight LDP). A randomized algorithm $\mathcal{M}$ provides ε-edge weight LDP, where $\varepsilon > 0$, if and only if for any two adjacency lists $\mathbf{a}$ and $\mathbf{a}'$ that only differ in one bit (called neighbors), and for any possible output $s \in Range(\mathcal{M})$,

$$\Pr\left[\mathcal{M}\left(\mathrm{a}\right) = s\right] \leq e^{\varepsilon} \Pr\left[\mathcal{M}\left(\mathrm{a}'\right) = s\right]. \tag{5}$$

Edge weight LDP is the same as edge LDP [35] except that the former (resp. latter) considers weighted graph (resp. unweighted graph). Under this definition, two adjacency lists differ by exactly one bit, indicating that they can differ on an edge and can also differ on the weight of an edge. Therefore, if a mechanism satisfies edge weight LDP, the impact of any edges or any edge weights on the final output is bounded, which guarantees the privacy of edges and edge weights.

2.3 Problem Statement

In our paper, we are interested in weighted graph data with temporal attribute. In particular, consider a decentralized network system where a data curator periodically collects information from n users. Based on the collected data at each time step, the curator sequentially publishes weighted graph snapshots at these time steps. Let $G_t = (V, E_t, a_t)$ be a undirected weighted graph snapshot at time steps t. $V = \{v_1, v_2, \cdots, v_n\}$ is the set of all n users. $E_t \subseteq V \times V$ is the set of edges, where an edge $(v_i, v_j) \in E_t$ denotes a relationship between users v_i and v_j at time step t. The weight function $a_t : E_t \longrightarrow \mathbb{R}$ maps edge (v_i, v_j) at time step t to a real weight. For each node v_i in G_t, the degree $(d_t)_i$ denotes the number of edges connected to the node and $(A_t)_i = (a_t(v_i, v_1), \cdots, a_t(v_i, v_n))$ denotes its adjacency list, where $a_t(v_i, v_j) \geq 1$ if $(v_i, v_j) \in E_t$, and otherwise $a_t(v_i, v_j) = 0$. The adjacency lists of all nodes at time step t form the adjacency matrix of graph G_t, formalized as $A_t = \{(A_t)_1, (A_t)_2, \cdots, (A_t)_n\}$. Since the data curator is untrustworthy, our goal is to design a LDP solution that helps the curator to collect user information and then construct a synthetic weighted graph snapshot $\hat{G}_t$ at each time step t. The resulting synthetic snapshot sequence $\hat{\mathcal{G}} = \left\langle \hat{G}_1, \hat{G}_2, \cdots \right\rangle$ is representative, i.e., it can support any downstream graph statistical analysis task while preserving individual privacy.

To balance utility loss and privacy loss over infinite sequences, we follow w-event-level privacy model [23] and extend its definition to the local setting for weighted graphs. Before that, we first introduce some notions. We call two adjacency lists $(A_t)_i$, $(A_t)'_i$ at time step t are neighboring if they only differ in one bit. Let a time-series $\mathcal{X} = \langle (A_1)_i, (A_2)_i, \cdots \rangle$, the definition of w-neighboring time-series is described as follows:

Definition 2.3 (w-neighboring time-series). For a positive integer w, two time-series $\mathcal{X}, \mathcal{X}'$ of length T are w-neighboring, if for each $\mathcal{X}[t], \mathcal{X}'[t]$ such that $t \in [T]$ and $\mathcal{X}[t] \neq \mathcal{X}'[t]$, it holds that $\mathcal{X}[t]$, $\mathcal{X}'[t]$ are neighboring; and for each $\mathcal{X}[t_1], \mathcal{X}[t_2], \mathcal{X}'[t_1], \mathcal{X}'[t_2]$ with $t_1 < t_2$, $\mathcal{X}[t_1] \neq \mathcal{X}'[t_1]$ and $\mathcal{X}[t_2] \neq \mathcal{X}'[t_2]$, it holds that $t_2 - t_1 + 1 \leq w$.

That is to say, if $\mathcal{X}, \mathcal{X}'$ are w-neighboring time-series, then their elements are the same or neighboring, and all their neighboring elements can fit in a window of at most w time steps. Hence, we define w-event edge weight LDP below.

Definition 2.4 (w-event edge weight LDP). Let $\mathcal{M}$ be a randomized algorithm that takes as input time-series $\mathcal{X}$ consisting of a single user's consecutive adjacency lists. We say that $\mathcal{M}$ provides w-event ε-edge weight LDP if and only if for any w-neighboring time-series $\mathcal{X}, \mathcal{X}'$, and for any possible output $S \in Range(\mathcal{M})$,

$$\Pr\left[\mathcal{M}\left(\mathcal{X}\right) = S\right] \leq e^{\varepsilon} \Pr\left[\mathcal{M}\left(\mathcal{X}'\right) = S\right]. \tag{6}$$

Discussion. Definition 2.4 aims to guarantee each user ε-edge weight LDP for any sliding window including w consecutive time steps, where ε can be regarded as the total available privacy budget in this sliding window. In other words, w-event edge weight

[1] Please refer to [27] for details of the EMS algorithm.

LDP ensures that the impact of the user's events in any w consecutive graph snapshots on the query result is limited. Therefore, an attacker cannot infer individual's information at any w consecutive time steps by observing the final sequence of synthetic graph snapshots. Note that when $w = 1$, the privacy protection level degenerates to event-level.

Example. We assume that w is 2. If the time-series of node v_1 with length 4 is $\mathcal{X} = \langle(0,1,3),(0,0,6),(0,2,3),(0,2,1)\rangle$, then one of its w-neighboring time-series can be $\mathcal{X}' = \langle(0,1,3),(0,1,6),(0,2,5),(0,2,1)\rangle$. This is because elements $(0,1,6)$ and $(0,2,5)$ in $\mathcal{X}'$ (i.e., adjacency lists) are neighboring to elements $(0,0,6)$ and $(0,2,3)$ in $\mathcal{X}$, respectively, and these neighboring elements are within the time window of size 2. The w-event edge weight LDP makes the probability that an attacker can distinguish between $\mathcal{X}$ and $\mathcal{X}'$ by observing the output of the randomized algorithm $\mathcal{M}$ controlled by the privacy budget ε.

3 OUR APPROACH

3.1 Overview

To publish weighted graphs at each time step with w-event edge weight LDP, we propose a new framework WGT-LDP. Figure 2 shows the workflow of WGT-LDP, which consists of the following four phases:

- **Population Division-based Sampling.** This component aims to sample nodes with large data changes at each time step. We adopt a population division-based approach, where the sampling nodes at the current time step have never been sampled in the previous $w-1$ time steps. To this end, the curator calculates the number of nodes whose noisy change error of the adjacency list are greater than a threshold δ among the current remaining nodes and adaptively adjusts the node sampling ratio. Then, the curator focuses on the information in the subgraph composed of all sampling nodes.
- **Data Range Estimation.** This component aims to estimate the local range of each sampling node's data in the subgraph to reduce the impact of perturbation noise on edge weights. In particular, each sampling node sends the perturbed local maximum weight to the curator, who adopts Expectation Maximization with Smoothing (EMS) algorithm and Bayesian estimation to determine their noisy maximum weights.
- **Aggregate Information Collection.** This component aims to collect the privatized degrees and adjacency lists of the sampling nodes in the subgraph. For each sampling node, it independently perturbs its degree and adjacency list based on the estimated local data range, and then sends them to curator. The curator performs post-processing to the collected perturbation information.
- **Graph Snapshot Generation.** This component aims to generate a weighted graph snapshot at the current time step. For the edges and weights between sampling nodes, the curator first adopts the noisy degree sequence as a characterization of the graph structure and the values in the noisy adjacency lists as a reference to rebuild the edges. Then the EMS algorithm and Bayesian estimation are applied again to rebuild the weights on these edges. For the edges and weights between non-sampling nodes, the curator uses the data of the previous time step to rebuild them.

Algorithm 2 Overall protocol of WGT-LDP framework

Input: Node set V of size n, original adjacency matrices $\langle A_1, A_2, \cdots, A_t, \cdots\rangle$, total privacy budget $\varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4$, window size w, threshold δ, transformation probabilities M, bucket size r
Output: Synthetic snapshots $\hat{\mathcal{G}} = \left\langle \hat{G}_1, \hat{G}_2, \cdots, \hat{G}_t, \cdots \right\rangle$ for publication

Initialize $\hat{G}_0$ with adjacency matrix $A_0 \in 0^{n\times n}$;
for each time step t **do**
  Obtain remaining node set $V_t = V \setminus \sum_{i=t-w+1}^{t-1} V_s^t$;
  $V_s^t = Sampling(\hat{G}_{t-1}, V_t, A_t, \delta, w, \varepsilon_1)$; ▷ Algorithm 2
  for each node $v_i^t \in V_s^t$ **do**
    Construct sub-adjacency list $(A_s^t)_i$ involving V_s^t;
    $h_i^t = \max((A_s^t)_i)$;
    $(d_s^t)_i = |\{j \mid (A_s^t)_{ij} \neq 0\}|$;
  end for
  $\hat{h}_t = Estimation((h_1^t, \cdots h_{s_t}^t), M, r, \varepsilon_2)$; ▷ Algorithm 3
  $\tilde{A}_s^t, \hat{d}_s^t = Collection(d_s^t, A_s^t, \hat{h}_t, \varepsilon_3, \varepsilon_4)$; ▷ Algorithm 4
  $\hat{G}_t = Generation(\hat{G}_{t-1}, \tilde{A}_s^t, \hat{d}_s^t, \hat{h}_t, M, r)$; ▷ Algorithm 5
end for
return $\hat{\mathcal{G}}$

Algorithm 2 gives the complete synthesis process of WGT-LDP. In the following, we will present each component in detail.

3.2 Population Division-based Sampling

For continuous publication of infinite sequences with w-event-level privacy, a general approach is to allocate privacy budgets for any sliding window containing w time steps. However, in our scenario, multiple components for publishing graph snapshots have to consume privacy budget, so the privacy budget of the sampling time steps also needs to be redistributed. This may lead to high noise level that destroy the utility of the synthetic graph. LDP-IDS [36] proposed to assign users to multiple groups, each of which uses the entire privacy budget in a window. This method provides w-event-level privacy under parallel composition. Unfortunately, random user sampling scheme ignores the differences in user data changes, which will introduce bias to the generation of graph data.

To strike the trade-off between high noise and information bias, we design a new population division-based sampling method that considers data changes. Specifically, a portion of the entire privacy budget ε is divided evenly to each time step in the time window for estimating the data dynamics. After that, the curator decides which nodes need to be sampled at each time step based on the number of remaining nodes and the noisy data change error of these nodes. Subsequently, this group of sampling nodes will use another portion of the privacy budget to perturb the local data between them. Since any node is only allowed to be sampled once in each sliding window, further allocation of the perturbation budget in the window is avoided, and the cost of each node does not exceed the ε-edge weight LDP budget.

Let V_t be the remaining node set at time step t, which is calculated by removing the already sampled nodes in the previous $w-1$ time steps from the node set V. For each node v_i in V_t with

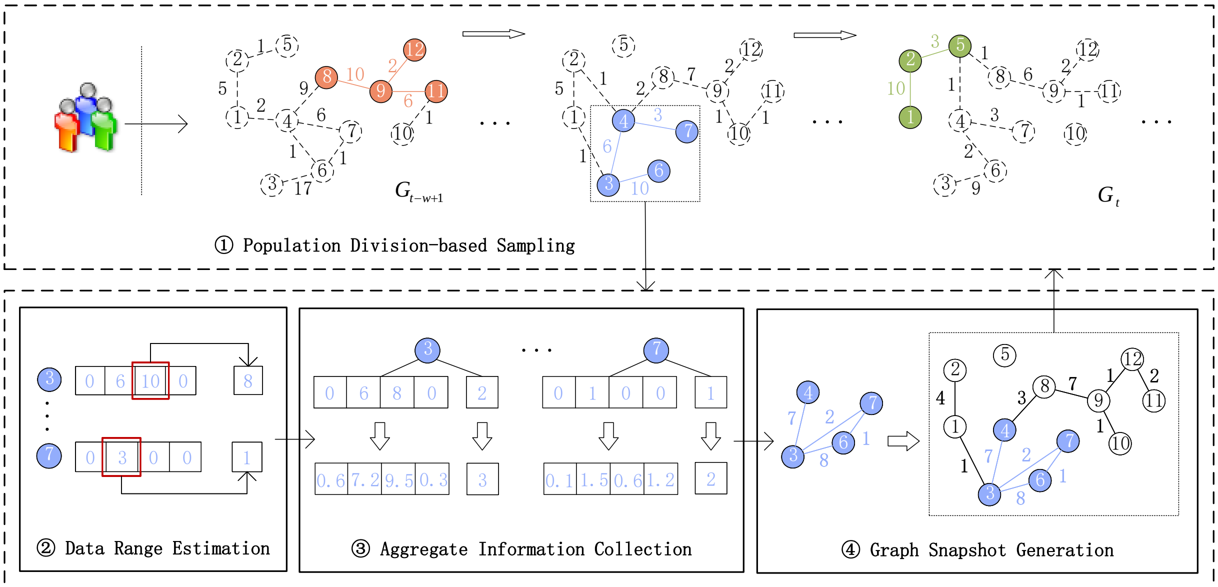


Figure 2: The overview of our WGT-LDP framework.

Algorithm 3 Population Division-based Sampling

Input: Published weighted graph $\hat{G}_{t-1}$, remaining node set V_t, adjacency list $(A_t)_i \in R^n$ of node $v_i \in V$, threshold δ, window size w, privacy budget ε_1

Output: Sampling node set V_s^t

// ***Collect change errors***

for each node $v_i \in V_t$ **do**

  $E_i^t = |\{j \mid a_t(v_j, v_i) \neq \hat{a}_{t-1}(v_j, v_i), v_j \in V_t\}|$;

  $\tilde{E}_i^t = GM(E_i^t, \frac{\varepsilon_1}{w}, \Delta f_{err})$;

end for

// ***Sampling strategy***

$m_t = |\{i \mid \tilde{E}_i^t \geq \delta, v_i \in V_t\}|$;

$P_t = 1 - \exp(-\frac{n}{w \cdot m_t})$;

$s_t = m_t \cdot P_t$;

Sort V_t in descending order of $\tilde{E}_i^t$;

Select top s_t in V_t as sampling node set V_s^t;

return V_s^t

adjacency list $(A_t)_i$, we first define its data change error E_i^t as the number of unequal bits involving V_t between adjacency lists $(A_t)_i$ and $(\hat{A}_{t-1})_i$, where $(\hat{A}_{t-1})_i$ is the noisy adjacency list of v_i in the previous synthetic graph $\hat{G}_{t-1}$. That is,

$$E_i^t = \left|\{j \mid a_t(v_j, v_i) \neq \hat{a}_{t-1}(v_j, v_i), v_j \in V_t\}\right|. \tag{7}$$

When there is no previous synthetic graph, i.e., at the first time step, we directly adopt $E_i^t = \left|\{j \mid a_t(v_j, v_i) \neq 0, v_j \in V_t\}\right|$ as the change error. In other words, we assume that the edges and weights in the previous synthetic graph are all 0. Then, We let each node locally add Geometric noise to E_i^t with privacy budget $\frac{\varepsilon_1}{w}$ to ensure w-event edge weight LDP, and send the privatized error $\tilde{E}_i^t$ to the curator. Note that the global sensitivity Δf_{err} of this query is 1 because changing one bit in the adjacency list result in the change of E_i^t by at most 1.

After receiving the noisy errors from all remaining nodes, the curator calculates the number of nodes whose noisy errors are greater than a threshold δ:

$$m_t = \left|\{i \mid \tilde{E}_i^t \geq \delta, v_i \in V_t\}\right|. \tag{8}$$

A simple solution is to sample all these m_t nodes. However, too large m_t will result in very few nodes that can be sampled at the next $w - 1$ time steps since any node only participates once in a window. In other words, the data between the large number of non-sampling nodes in the subsequent graph need to be approximated by the previous time step, which may cause excessive approximation deviation for graphs with significant dynamic changes. Therefore, the nodes that can be sampled at each time step in the sliding window are limited and should be carefully allocated. To this end, the curator adaptively adjusts the node sampling ratio based on m_t. We choose $\frac{n}{w}$ as the adjustment threshold since it implies the theoretical average number of sampling nodes at each time step in the window. In particular, when m_t is less than $\frac{n}{w}$, the m_t nodes should be sampled as much as possible to effectively capture data changes. While when m_t is greater than $\frac{n}{w}$, the number of samples should be reduced to ensure that there are still enough nodes available for sampling in future time steps, which can maintain the utility of the synthetic graph. Based on the above analysis, we set the sampling ratio as follows:

$$P_t = 1 - \exp(-\frac{n}{w \cdot m_t}). \tag{9}$$

The final number of sampled nodes obtained at the current time step is calculated as $s_t = m_t \cdot P_t$, and the corresponding set of sampling node, denoted as $V_s^t = \{v_i^t \mid i \in [s_t]\}$, consists of the nodes with the top s_t data change errors among the m_t nodes. This process is described in Algorithm 3.

Algorithm 4 Data Range Estimation

Input: Maximum value h_i^t in sub-adjacency list of node $v_i^t \in V_s^t$, transformation probabilities M, bucket size r, privacy budget ε_2
Output: Noisy maximum values $\hat{h}_t = \{\hat{h}_1^t, \cdots, \hat{h}_{s_t}^t\}$

// Collect values
for each node $v_i^t \in V_s^t$ **do**
  $x_i^t = h_i^t / B$;
  $\tilde{x}_i^t = SW(x_i^t, \varepsilon_2)$;
end for
// Post-processing
$\tilde{\mathbf{x}} = \{\tilde{x}_1^t, \cdots, \tilde{x}_{s_t}^t\}$;
$\tilde{\mathbf{z}} = EMS(\tilde{\mathbf{x}}, M, r)$;
for $i \in \{1, 2, \cdots, s_t\}$ **do**
  Calculate $\hat{x}_i^t$ based on $\tilde{\mathbf{z}}$ by Eq. 10 and 11;
  $\hat{h}_i^t = B \cdot \hat{x}_i^t$;
end for
return $\hat{h}_t$

3.3 Data Range Estimation

Given the sampling nodes at time step t, the curator focuses on the data of the subgraph G_s^t composed of these nodes, which has a public data-independent upper bound B, i.e., the maximum possible weight in G_s^t. To satisfy edge weight LDP, the noise added to the local edge weight of each node needs to be proportional to B, which may lead to large edge weight errors. If users are only required to perturb the edge weights based on their true local data domain, differential privacy cannot be satisfied, because the upper bound of the domain itself is private information. In the paper, we propose a method to estimate the local data domain of each sampling node.

In particular, let the sub-adjacency matrix of the subgraph G_s^t be A_s^t, where row vector $(A_s^t)_i$ is the sub-adjacency list of the sampling node v_i^t. The local data range of node v_i^t is $[0, h_i^t]$, where $h_i^t \in [0, B]$ is the maximum value in the sub-adjacency list of the node. Given the privacy budget ε_2 and parameter $b = \frac{\varepsilon_2 e^{\varepsilon_2} - e^{\varepsilon_2} + 1}{2e^{\varepsilon_2}(e^{\varepsilon_2} - 1 - \varepsilon_2)}$, each sampling node transforms h_i^t to $x_i^t = \frac{h_i^t}{B} \in [0, 1]$, and uses the Square Wave mechanism (SW) to report a value close to x_i^t with probability $p = \frac{e^{\varepsilon_2}}{2be^{\varepsilon_2}+1}$, which carries useful information about x_i^t, and then sends the perturbed value $\tilde{x}_i^t$ to the curator.

In order to improve the estimation accuracy and reduce the bias, we perform a post-processing step on the perturbed values. First, we use Expectation Maximization with smoothing (EMS) algorithm [27] (described in Algorithm 1) to infer the probability distribution $\tilde{\mathbf{z}}$ of the original value x_i^t of all sampling nodes. To run EMS, the perturbed value $\tilde{x}_i^t$ reported by the user need to be discretized into r buckets in output domain $[-b, 1 + b]$. In this phase, we set the number of buckets to $r = \lfloor B \rfloor$.

After that, the probability distribution $\tilde{\mathbf{z}}$ is used as a prior, such that the curator can apply Bayes' theorem to calculate the corresponding posterior probability distribution, i.e., for each $i \in [s_t]$,

$$P(x_i^t \in b_k \mid \tilde{x}_i^t \in b_j) = \frac{M_{j,k} \cdot \tilde{\mathbf{z}}_k}{\sum_{k=1}^{r} M_{j,k} \cdot \tilde{\mathbf{z}}_k}. \tag{10}$$

Algorithm 5 Aggregate Information Collection

Input: Degree $(d_s^t)_i$, sub-adjacency list $(A_s^t)_i$ and noisy maximum values $\hat{h}_t = \{\hat{h}_1^t, \cdots, \hat{h}_{s_t}^t\}$, privacy budget $\varepsilon_3, \varepsilon_4$
Output: Noisy degree sequence $\hat{d}_s^t$, noisy sub-adjacency matrix $\tilde{A}_s^t$

// Collect degrees and sub-adjacency lists
for each node $v_i^t \in V_s^t$ **do**
  $(\tilde{d}_s^t)_i = GM((d_s^t)_i, \varepsilon_3, \Delta f_d)$;
  $a_t(v_i^t, v_j^t) = Truncate(a_t(v_i^t, v_j^t), \hat{h}_i^t)$;
  $b_t(v_i^t, v_j^t) = a_t(v_i^t, v_j^t) / \hat{h}_i^t$;
  $\tilde{b}_t(v_i^t, v_j^t) = SW(b_t(v_i^t, v_j^t), \varepsilon_4)$;
end for
// Degrees adjustment
$\hat{d}_s^t = NormSub(\tilde{d}_s^t)$;
if $\sum_{i=1}^{s_t} (\hat{d}_s^t)_i$ is odd **then**
  randomly select a node degree $(\hat{d}_s^t)_i$;
  flip a coin to decide whether to add 1 or subtract 1;
end if
return $\tilde{A}_s^t, \hat{d}_s^t$

For each sampling node v_i^t, the curator selects the upper bound of the bucket with the maximum posterior probability as its noisy value, i.e.,

$$\hat{x}_i^t = \sup(\arg\max_{b_k} P(x_i^t \in b_k \mid \tilde{x}_i^t \in b_j)), \tag{11}$$

and calculates the corresponding noisy local maximum weight as $\hat{h}_i^t = B \cdot \hat{x}_i^t$. Algorithm 4 shows the estimation procedure.

3.4 Aggregate Information Collection

Based on the estimated data range, the sampling nodes locally perturb important information about the topology of the subgraph G_s^t at time step t, including node degrees and sub-adjacency lists. In particular, the node degrees reflect the subgraph density and have low sensitivity, which can better denoise the subgraph topology. The sub-adjacency lists contain fine-grained information of weighted connections between sampling nodes, which can be used as a reference for edge generation and reconstruct edge weights.

Let $(d_s^t)_i$ be the degree of sampling node v_i^t in the subgraph G_s^t. As shown in Algorithm 5, given the privacy budget ε_3, each sampling node v_i^t use Geometric mechanism to inject unbiased noise into its degree. The sensitivity of degree Δf_d is 1 because adjusting one bit from the adjacency list of a node changes its degree by at most 1. For sub-adjacency list $(A_s^t)_i$, the sampling node v_i^t maps each bit $a_t(v_i^t, v_j^t)$ to $b_t(v_i^t, v_j^t) \in [0, 1]$ according to the estimated data range $[0, \hat{h}_i^t]$, and then uses the SW to randomly perturb these bit with the privacy budget ε_4. Note that when the value of a bit is larger than $\hat{h}_i^t$, it will first be truncated, i.e., $a_t(v_i^t, v_j^t) = \min(a_t(v_i^t, v_j^t), \hat{h}_i^t)$. Since the SW exploits the ordinal property of edge weights, a report that is different from but close to the true weight also carries useful information about the weight, which is exactly what we expect.

After adding Geometric noise, some perturbed degree values may appear negative and their sum may be odd, which makes the subsequent graph generation infeasible. Thus, we first adopt

Algorithm 6 Graph Snapshot Generation

Input: Published weighted graph $\tilde{G}_{t-1}$, noisy degree sequence $\hat{d}_s^t$, noisy sub-adjacency matrix $\tilde{A}_s^t = (\tilde{b}_t(v_i^t, v_j^t))$, noisy maximum values $\hat{h}_t = \{\hat{h}_1^t, \cdots, \hat{h}_{s_t}^t\}$, transformation probabilities M, bucket size r

Output: Synthetic weighted graph $\hat{G}_t$ for publication

```
Initialize Ĝ_t = (Ĝ_s^t, Ĝ_t\Ĝ_s^t) as a null graph;
// Generate edges and weights in Ĝ_s^t
while d̂_s^t ≠ 0 or |d̂_s^t > 0| ≠ 1 do
    Choose a node v_i^t with (d̂_s^t)_i is minimal positive entry in d̂_s^t;
    for each v_j^t ≠ v_i^t with (d̂_s^t)_j is the positive entry do
        Pick v_j^t with j = arg max_j b̃_t(v_i^t, v_j^t) · b̃_t(v_j^t, v_i^t);
    end for
    Add edge (v_i^t, v_j^t) to Ĝ_s^t;
    (d̂_s^t)_i = (d̂_s^t)_i − 1, (d̂_s^t)_j = (d̂_s^t)_j − 1;
end while
Construct sequence b̃ by bits corresponding to all edges in Ĝ_s^t;
z̃ = EMS(b̃, M, r);
for each b̃_t(v_i^t, v_j^t) ∈ b̃ do
    Calculate b̂_t(v_i^t, v_j^t) based on z̃ by Eq. 10 and 11;
    â_t(v_i^t, v_j^t) = ĥ_i^t · b̂_t(v_i^t, v_j^t);
    Add weight â_t(v_i^t, v_j^t) to Ĝ_s^t;
end for
// Generate edges and weights in Ĝ_t\Ĝ_s^t
Approximate with the corresponding values Ĝ_{t−1}
return Ĝ_t
```

NormSub [42] to post-process the negative value problem. Given perturbed degree sequence $\tilde{d}_s^t = \langle (\tilde{d}_s^t)_1, (\tilde{d}_s^t)_2, \cdots, (\tilde{d}_s^t)_{s_t} \rangle$, our goal is to find an optimal integer $\alpha^* = \arg\min_{\alpha} |\sum_{i \in [s_t]} \max((\tilde{d}_s^t)_i + \alpha, 0) - \sum_{i \in [s_t]} (\tilde{d}_s^t)_i|$. After obtaining α^*, each node degree $(\tilde{d}_s^t)_i$ is updated to $(\hat{d}_s^t)_i = \max((\tilde{d}_s^t)_i + \alpha^*, 0)$, which satisfies non-negative constraint. To solve the odd sum problem, we randomly select a node degree in $\hat{d}_s^t$ and flip a coin to decide whether to add 1 or subtract 1 to this node degree.

3.5 Graph Snapshot Generation

Now, the curator generates the synthetic subgraph $\hat{G}_s^t$ at time step t based on the noisy degree sequence $\hat{d}_s^t$ and the noisy sub-adjacency matrix $\tilde{A}_s^t$ (consisting of the noisy sub-adjacency lists from all sampling nodes). We first use $\hat{d}_s^t$ to characterize the graph structure and use $\tilde{A}_s^t$ as a reference to reconstruct the edges in $\hat{G}_s^t$. The intuition of this method is that the bits with high noisy values in the sub-adjacency matrix are more likely to correspond to non-zero edges in the original subgraph.

As shown in Algorithm 6, the curator first picks a node v_i^t with the minimum degree and considers other nodes with non-zero degree in $\hat{d}_s^t$ as candidate nodes that may be connected to v_i^t. For potential edges between v_i^t and each candidate node v_j^t, the curator refers to the bits $\tilde{b}_t(v_i^t, v_j^t)$ and $\tilde{b}_t(v_j^t, v_i^t)$ in $\tilde{A}_s^t$ related to their existence and adds a potential edge with the largest bit product to $\hat{G}_s^t$. That is, the index of the node v_j^t connected to v_i^t is determined to be

$$j = \arg\max_j \tilde{b}_t(v_i^t, v_j^t) \cdot \tilde{b}_t(v_j^t, v_i^t). \tag{12}$$

Then, the degrees corresponding to nodes v_i^t and v_j^t in $\hat{d}_s^t$ are both subtracted by 1. The above process will be repeated until $\hat{d}_s^t$ is reduced to 0 or no additional edges can be added.

To reconstruct the weights on the edges, the curator considers the bit $\tilde{b}_t(v_i^t, v_j^t)$ corresponding to each edge (v_i^t, v_j^t) in $\hat{G}_s^t$, and then post-processes $\tilde{b}_t(v_i^t, v_j^t)$ by applying the EMS algorithm and Bayesian estimation described in Section 3.3. In this phase, we set the number of buckets to $r = \lfloor \max(\hat{h}_i^t) \rfloor$. After obtaining the post-processed noisy bit $\hat{b}_t(v_i^t, v_j^t)$, the noisy weight on the edge (v_i^t, v_j^t) is calculated as $\hat{a}_t(v_i^t, v_j^t) = \hat{h}_i^t \cdot \hat{b}_t(v_i^t, v_j^t)$.

For the edges and weights of non-sampling nodes, they approximate the values in the last generated graph. At this point, the synthetic weighted graph snapshot $\hat{G}_t$ at time step t is generated.

4 THEORETICAL ANALYSIS

4.1 Privacy Analysis

In this part, we prove that WGT-LDP achieves w-event ε-edge weight LDP. Before that, we first give the following lemmas to show that the components of WGT-LDP satisfy w-event edge weight LDP. To facilitate the proof, we write the adjacency list $(A_t)_i$ at time step t as $\mathbf{a}_t$.

Lemma 4.1. *Population Division-based Sampling satisfies w-event ε_1-edge weight LDP.*

Proof. In the phase, the remaining nodes at each time step add Geometric noise to their data change errors with privacy budget $\frac{\varepsilon_1}{w}$. Considering the independent randomness of all mechanisms, for time-series $\mathcal{X}$ and an arbitrary vector $S = (s_1, s_2, \cdots, s_T) \in Range(\mathcal{M})$, we have

$$\Pr[\mathcal{M}(\mathcal{X}) = S] = \prod_{t=1}^{T} \Pr[\mathcal{M}_t(\mathbf{a}_t) = s_t]. \tag{13}$$

Since the Geometric mechanism can provide strict privacy guarantees (review 2.1), there is $\frac{\Pr[\mathcal{M}_t(\mathbf{a}_t)=s_t]}{\Pr[\mathcal{M}_t(\mathbf{a}'_t)=s_t]} \leq e^{\frac{\varepsilon_1}{w}}$, where $\mathbf{a}_t$ and $\mathbf{a}'_t$ are neighboring.

By Definition 2.3, there exists $i \in [T]$, such that $\mathbf{a}_t = \mathbf{a}'_t$ for $1 \leq t \leq i - w$ and $i + 1 \leq t \leq T$. Moreover, a node needs to report the error at most w times as a time window consisting of w time steps. Thus, for any w-neighboring time-series $\mathcal{X}, \mathcal{X}'$, we have

$$\frac{\Pr[\mathcal{M}(\mathcal{X}) = S]}{\Pr[\mathcal{M}(\mathcal{X}') = S]} \leq \prod_{t=i-w+1}^{i} \frac{\Pr[\mathcal{M}_t(\mathbf{a}_t) = s_t]}{\Pr[\mathcal{M}_t(\mathbf{a}'_t) = s_t]} \tag{14}$$

$$\leq e^{\sum_{t=i-w+1}^{i} \frac{\varepsilon_1}{w}} \tag{15}$$

$$= e^{\varepsilon_1} \tag{16}$$

The subsequent node sampling process does not touch the true adjacency lists and does not consume privacy budget. Therefore, population division-based sampling satisfies w-event ε_1-edge weight LDP. □

Lemma 4.2. *Data Range Estimation satisfies w-event ε_2-edge weight LDP.*

Proof. We prove the lemma by proving the following two claims:

(1) In any sliding window containing w time steps, each node is sampled by the curator at most once.
(2) Each node's reported data satisfies ε_2-edge weight LDP.

First, at each time step t, since the sampling nodes V_s^t with size s_t always come from the remaining nodes V_t calculated in the previous stage, we can get that the total number of sampling nodes in a time window of size w is not larger than n, i.e., $\sum_{t=i-w+1}^{i} s_t \leq n$ for $i \in [T]$. Moreover, we have $V_s^1 \cap V_s^2 \cap \cdots \cap V_s^n = \emptyset$. Therefore, the first claim holds.

Then, we prove the second claim. Given two adjacency lists $\mathbf{a}_t$ and $\mathbf{a}_t'$ differ at exactly one bit, let x_t and x_t' be their corresponding maximum bits after normalization. For any possible output $s \in Range(\mathcal{M})$, combining the probability equation of Square Wave mechanism, we have

$$\frac{\Pr\left[\mathcal{M}\left(\mathbf{a}_t\right)=s\right]}{\Pr\left[\mathcal{M}\left(\mathbf{a}_t'\right)=s\right]}=\frac{\Pr\left[\tilde{x}_t=s\right]}{\Pr\left[\tilde{x}_t'=s\right]} \leq \frac{e^{\varepsilon_2}\cdot q}{q}=e^{\varepsilon_2} \tag{17}$$

The post-processing process does not lose privacy. To sum up, data range estimation satisfies w-event ε_2-edge weight LDP. □

Lemma 4.3. *Aggregate Information Collection satisfies w-event $(\varepsilon_3+\varepsilon_4)$-edge weight LDP.*

Proof. In this phase, the sampling nodes perturb their degrees and adjacency lists using Geometric mechanism and Square Wave mechanism respectively. Since these sampling nodes only report their data once in a window of size w, according to Lemma 4.2, we only need to prove that the degree (resp. adjacency list) reported by each node satisfy ε_3 (resp. ε_4)-edge weight LDP. Recalling 2.1, Geometric mechanism and Square Wave mechanism can provide rigorous privacy guarantee. Based on sequential composition property, aggregate information collection satisfies w-event $(\varepsilon_3+\varepsilon_4)$-edge weight LDP. □

Theorem 4.4. *WGT-LDP satisfies w-event ε-edge weight LDP, where $\varepsilon=\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4$.*

Proof. According to the above lemma, the first three components of WGT-LDP satisfy ε_1, ε_2, $\varepsilon_3+\varepsilon_4$-edge weight LDP, respectively. In graph snapshot generation, WGT-LDP processes the perturbed data without consuming privacy budget. Therefore, WGT-LDP satisfies w-event ε-edge weight LDP with sequential composition. □

4.2 Utility Analysis

In this part, we theoretically justify the utility guarantee of WGT-LDP. First, we need to determine a metric that can reflect the quality of the synthetic weighted graph $\hat{G}_t$ at each time step t, which can be used to quantify the utility of WGT-LDP. Many previous literature on differentially private graph analysis uses metrics that only measure the utility of specific graph analysis tasks, such as subgraph counts [18], degree distribution [16], and clustering coefficient [47], which cannot directly reflect the quality of the synthetic graph. In contrast, the performance of the synthetic graph depends on whether it is semantically similar to the original graph, so that it can support any downstream graph statistical analysis tasks. Since different tasks have different sensitivities to graph structural properties, e.g., the clustering coefficient depends on the local connection pattern of the graph and the path condition depends on the connectivity of the graph, the synthetic graph needs to find a feasible solution between high-dimensional structural perturbations and multi-objective statistical utility preservation. This is one of the reasons why many previous studies on differentially private synthetic graphs [2, 29, 35, 48] only provide empirical evidence on utility. Nevertheless, one intuition is that the closer the adjacency matrix $\hat{A}_t$ of synthetic weighted graph $\hat{G}_t$ is to the adjacency matrix A_t of original weighted graph G_t, the better the performance of $\hat{G}_t$ on downstream tasks. Therefore, inspired by the utility analysis of GNN performance in Blink [50], we use the expected $\ell1$-distance between $\hat{A}_t$ and A_t at any time step t, i.e., $E[\|\hat{A}_t-A_t\|_{1,1}]$, to evaluate the utility of WGT-LDP, which can reflect the overall quality of any synthetic weighted graph snapshot G_t generated under w-event edge weight LDP.

Unlike the error estimation of tabular stream data with w-event LDP [36], the expected $\ell1$-distance of WGT-LDP is complicated. To simplify the utility analysis, we assume that the edge weights between non-sampling nodes are approximated by 0. Besides, for the subgraph G_s^t composed of sampling nodes, its maximum edge weight h_{max}^t is at most $\hat{h}_{max}^t=\max\{\hat{h}_1^t,\cdots,\hat{h}_{s_t}^t\}$, where $\hat{h}_i^t$ represents the noisy maximum value in sub-adjacency list of sampling node v_i^t. Then WGT-LDP has the following utility guarantee.

Theorem 4.5. *Assume that the edge weights between non-sampling nodes are approximated by 0. Then, for any time step $t \in \mathbb{Z}_{\geq 0}$, $\hat{h}_{max}^t \in \mathbb{R}_{\geq 0}$, and subgraph $G_s^t \in G_t$ composed of sampling nodes such that the maximum edge weight h_{max}^t of G_s^t is at most $\hat{h}_{max}^t$, we have*

$$E[\|\hat{A}_t-A_t\|_{1,1}] \leq \frac{\hat{h}_{max}^t(b+1)}{2}\|A_t\|_{0,0}+\|A_t\|_{1,1}, \tag{18}$$

where $b=\frac{\varepsilon_4 e^{\varepsilon_4}-e^{\varepsilon_4}+1}{2e^{\varepsilon_4}(e^{\varepsilon_4}-1-\varepsilon_4)}$ and $\|A_t\|_{0,0}$ denotes the number of non-zero elements in A_t.

Proof. At each time step t, WGT-LDP first performs population division-based sampling to obtain a sampling node set $V_s^t=\left\{v_i^t \mid i \in [s_t]\right\}$. Due to the assumption that the edge weights between non-sampling nodes are approximated by 0. Therefore, we have

$$E[\|\hat{A}_t - A_t\|_{1,1}] = \mathrm{E}\left[\sum_{(i,j)\in[n]\times[n]} \left|(\hat{A}_t)_{ij} - (A_t)_{ij}\right|\right]$$
$$= \sum_{(i,j)\in[n]\times[n]} \mathrm{E}\left[\left|(\hat{A}_t)_{ij} - (A_t)_{ij}\right|\right]$$
$$= \sum_{(i,j)\in[s_t]\times[s_t]} \mathrm{E}\left[\left|(\hat{A}_s^t)_{ij} - (A_s^t)_{ij}\right|\right] + \sum_{(i,j)\in[n]\times[n]} \left|(A_t)_{ij}\right| - \sum_{(i,j)\in[s_t]\times[s_t]} \left|(A_s^t)_{ij}\right|. \tag{19}$$

Next, we analyze $\sum_{(i,j)\in[s_t]\times[s_t]} \mathrm{E}\left[\left|(\hat{A}_s^t)_{ij} - (A_s^t)_{ij}\right|\right]$, namely, the error introduced by generating synthetic subgraph $\hat{G}_s^t$. In the phase of data range estimation, we obtain the noisy upper bound $\hat{h}_i^t$ of the sub-adjacency list of each sampling node v_i^t. Then, the maximum edge weight h_{max}^t of G_s^t can use $\hat{h}_{max}^t = \max\{\hat{h}_1^t, \cdots, \hat{h}_{s_t}^t\}$ as the privacy estimate. When the maximum edge weight h_{max}^t of G_s^t is at most $\hat{h}_{max}^t$, the only randomness in the process of generating $\hat{G}_s^t$ will be the noise caused by the Square Wave mechanism (SW) and the Geometric mechanism (GM), since bias introduced by truncation in Algorithm 5 will not occur. For an input value $x \in [0, 1]$, given the privacy budget ε, the expected error of the output value y after SW as

$$E_{\mathrm{SW}}[|y - x|] = \int_{x-b}^{x+b} |y - x| \cdot p \, \mathrm{d}y + \int_{-b}^{x-b} (x - y) \cdot q \, \mathrm{d}y + \int_{x+b}^{1+b} (y - x) \cdot q \, \mathrm{d}y$$
$$= pb^2 + q(v^2 - v + b + \frac{1}{2})$$
$$\leq pb^2 + q\frac{1 + 2b}{2}. \tag{20}$$

Since $p = \frac{e^\varepsilon}{2be^\varepsilon + 1}$ and $q = \frac{1}{2be^\varepsilon + 1}$, we have

$$pb^2 + q\frac{1 + 2b}{2} = \frac{e^\varepsilon b^2 + b + \frac{1}{2}}{2be^\varepsilon + 1} \leq \frac{e^\varepsilon b^2 + b}{2be^\varepsilon} \leq \frac{b + 1}{2}. \tag{21}$$

Combining Eq. 20 and 21, we have the following

$$E_{\mathrm{SW}}[|y - x|] \leq \frac{b + 1}{2}, \tag{22}$$

where $b = \frac{\varepsilon e^\varepsilon - e^\varepsilon + 1}{2e^\varepsilon(e^\varepsilon - 1 - \varepsilon)}$.

In the phase of aggregate information collection, each bit $(A_s^t)_{ij}$ is perturbed with privacy budget ε_4 using SW and the degree $(d_s^t)_i$ of each sampling node is perturbed with privacy budget ε_3 using GM. Hence, for any $(A_s^t)_{ij} \in [0, \hat{h}_{max}^t]$,

$$\mathrm{E}\left[\left|(\hat{A}_s^t)_{ij} - (A_s^t)_{ij}\right|\right] \leq \hat{h}_{max}^t \frac{b + 1}{2}, \tag{23}$$

where $b = \frac{\varepsilon_4 e^{\varepsilon_4} - e^{\varepsilon_4} + 1}{2e^{\varepsilon_4}(e^{\varepsilon_4} - 1 - \varepsilon_4)}$. The expectation of the noisy degree $(\hat{d}_s^t)_i$ is $E[(\hat{d}_s^t)_i] = (d_s^t)_i$ since GM is unbiased.

Due to WGT-LDP generates $\hat{G}_s^t$ based on the noisy degree sequence in the final phase, the expected number of perturbed bits in $\hat{A}_s^t$ is at most $E(\sum_{i\in[s_t]} (d_s^t)_i) = \sum_{i\in[s_t]} E((d_s^t)_i) = \sum_{i\in[s_t]} (d_s^t)_i$. Then, we have

$$\sum_{(i,j)\in[s_t]\times[s_t]} \mathrm{E}\left[\left|(\hat{A}_s^t)_{ij} - (A_s^t)_{ij}\right|\right]$$
$$\leq \hat{h}_{max}^t \frac{b + 1}{2} \cdot \left(\sum_{i\in[s_t]} (d_s^t)_i\right) + \sum_{(i,j)\in[s_t]\times[s_t]\setminus S} \left|(A_s^t)_{ij}\right| \tag{24}$$
$$= \frac{\hat{h}_{max}^t(b + 1)}{2} \|A_s^t\|_{0,0} + \sum_{(i,j)\in[s_t]\times[s_t]\setminus S} \left|(A_s^t)_{ij}\right|,$$

where S is the set of index pairs corresponding to the perturbed bits.

Substitute Eq. 24 into Eq. 19, we have the final result

$$E[\|\hat{A}_t - A_t\|_{1,1}] \leq \frac{\hat{h}_{max}^t(b + 1)}{2} \|A_s^t\|_{0,0} + \sum_{(i,j)\in[s_t]\times[s_t]\setminus S} \left|(A_s^t)_{ij}\right| + \sum_{(i,j)\in[n]\times[n]} \left|(A_t)_{ij}\right| - \sum_{(i,j)\in[s_t]\times[s_t]} \left|(A_s^t)_{ij}\right|$$
$$\leq \frac{\hat{h}_{max}^t(b + 1)}{2} \|A_t\|_{0,0} + \|A_t\|_{1,1}. \tag{25}$$

□

Discussion. Theorem 4.5 shows that under w-event edge weight LDP, the expected $\ell 1$ error of WGT-LDP is positively correlated with the original graph G_t itself and its maximum edge weight. In fact, G_t is usually sparse, and only a small number of edges have large weights. In addition, the privacy budget ε_4 associated with the utility is independent of w, i.e., it does not need to be allocated to the sliding window of size w. Therefore, the synthetic graph $\hat{G}_t$ generated by WGT-LDP at any time step t is a reasonable representation of the original graph G_t. This benefits from the fact that the population division-based sampling increases the available privacy budget at each time step, the data range estimation reduces the upper bound of the maximum possible edge weight (from B to $\hat{h}_{max}^t$), and graph snapshot generation based on noisy degree sequence preserves the sparse structure of the graph.

4.3 Computational Complexity

Consider weighted graph publication with n nodes at time step t. For WGT-LDP, the process of population division-based sampling analyzes the data change errors based on the adjacency list of $|V_t|$ remaining nodes, which has time complexity is $O(|V_t|^2)$. The computational complexity of data range estimation is $O(r^2\rho + s_t)$, since it includes the post-processing step and then adjusts the data of s_t sampling nodes. The constants r and ρ are the number of buckets and the number of iterations in the EMS algorithm, respectively. In aggregate information collection, the s_t sampled nodes perturb their degrees and sub-adjacency lists. Hence, the computational complexity of aggregate information collection is $O(s_t^2)$. Since the phase of graph snapshot generation selects edges based on the degree sequence and performs post-processing on edge weights, the computational complexity of this phase is $O(r^2\rho + c_t)$, where c_t is the sum of the degrees of all sampling nodes. Above all, we have $O(|V_t|^2 + r^2\rho + s_t + s_t^2 + r^2\rho + c_t) < O(n^2 + r^2\rho)$. Thus, the total computational complexity of WGT-LDP is $O(n^2 + r^2\rho)$.

Table 1: Basic Information of Datasets

Datasets	n	T	B	Type
Email-Eu	319	173	72	communication
Forum	899	24	168	Social
Tech-AS	5000	24	24	Autonomous System
Synthetic-I & II	10000	100	200	Synthetic

5 EVALUATION

We conducted extensive experiments to evaluate the effectiveness of WGT-LDP. All experiments are conducted in Python on a laptop with Intel Core i5-1135G7 CPU, 16GB RAM. For each experiment, we performed it ten times and presented the average results.

5.1 Experimental Setup

Datasets. We first conduct experiments on three real-world graph datasets. Note that we regard the connection of nodes occurring within an hour is regarded as single connection.

- Email-Eu [34] contains 61046 dynamic email communications between 319 institution members. Each member on the network represents as a node. We use the number of communications between members to construct weighted topology of each graph snapshot.
- Forum [33] contains 33720 dynamic interaction records between 899 students in the community. We abstract each student as a node and construct weighted topology based on the number of interactions between students in a specific time window.
- Tech-AS [37] records 171403 dynamic connections between 34761 autonomous systems. We randomly selected 5000 autonomous systems from the network as nodes, and construct the weighted topology of each snapshot according to the number of connections between these nodes in a specific time window.

Secondly, we use NetworkX [14], an open source package in python, to generate two synthetic datasets that follow a power-law distribution, namely Synthetic-I and Synthetic-II. Both synthetic datasets contain 100 weighted graph snapshots with 10000 nodes, and the edge weight of each snapshot is bounded by 200. The difference is that the severity of snapshot changes between adjacent time steps in Synthetic-I is 20%, while that in Synthetic-II is 90%.

Table 1 summarizes the basic information of all datasets, where n is the numbers of nodes, T is the number of snapshots and B is the common upper bound of the edge weights.

Parameter Settings. In our experiments, we set the threshold $\delta = 1$ for node sampling. The sliding window size w is set to 5. For the allocation of privacy budget, we set $\varepsilon_1 = \frac{1}{2}\varepsilon$ and $\varepsilon_2 = \varepsilon_3 = \varepsilon_4 = \frac{1}{6}\varepsilon$, where the total privacy budget ε varies from 0.5 to 2.5.

Baselines. We compare WGT-LDP with the following baselines:

- **BDG.** This method is a baseline solution based on budget division, which does not perform node sampling but uniformly distributes the total privacy budget ε to w time steps in the sliding windows.
- **RPG.** The baseline RPG adopts the population division-based solution. However, it does not consider the data changes of nodes, i.e., it randomly assigns all nodes to w time steps in the window.
- **Swg-NS.** The mechanism SwgDP [46] considers the dynamic weighted graphs publishing mechanism with DP. Since the node adaptive sampling component in SwgDP can be easily modified to the LDP setting, we use this component to perform node sampling and synthesize weighted graphs in the subsequent phases. We denote this baseline as Swg-NS.
- **HMG.** Li et al. [29] proposed HMG as an LDP mechanism to achieve dynamic graph publishing for decentralized applications. Since HMG does not consider the edge weights, we add noisy weights perturbed by the Laplace mechanism [9] to each edge after HMG generates the edges in the weighted graph.

Note that to ensure the rationality of the comparison, the baselines Swg-NS and HMG are reproduced in the w-event-level privacy model.

In addition, we also implement two static unweighted graph mechanisms LDPGen [35] and Blink [50] for comparison, which can show the performance of our method on event-level privacy ($w = 1$). Since Blink is originally designed for training GNN models, we use their proposed variant Blink-Hard to generate synthetic graph at any time step. Both mechanisms are implemented under the privacy definition of edge LDP [35]. This definition can be viewed as the case where the value of each bit in Definition 2.2 is 0 or 1. Note that when $w = 1$, WGT-LDP does not sample users, but directly executes the last three phases to generate the entire weighted graph, which is equivalent to BDG and RPG.

Metrics. We evaluate the quality of the synthetic weighted graph snapshots from the following four aspects: degree distribution, weight distribution, clustering coefficient and path condition, where the first three are graph statistical queries and the last one is a graph structure query. Since these snapshots are continuously released, we calculate the average of each metric over all time steps. Note that for the clustering coefficient, we consider the Root Mean Squared Error (RMSE) over time. Smaller results indicate higher utility.

- **Degree Distribution.** We adopt Kullback-Leibler (KL) divergence [25] to evaluate the error of the degree distributions between the original and synthetic weighted graphs. To avoid the denominator in the KL divergence being zero, we add a small value to the degree distribution of the original graphs and the synthetic graphs.
- **Weight Distribution.** We bucketize the weights into B bins and count the number of edges that fall into each bin to calculate the weight distribution. Similar to the degree distribution, we use KL divergence to measure the error.
- **Clustering Coefficient.** It is a statistical metric that characterizes the community structure of a graph. We use the RMSE of the clustering coefficient over all time steps to measure the error.

$$\mathrm{RMSE}_{CC} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y - \tilde{Y}\right)^2}, \tag{26}$$

where Y and $\tilde{Y}$ denote the clustering coefficients of the original and synthetic weighted graphs, respectively.

- **Path Condition.** The path length of a graph measures its connectivity, which is denoted as the maximum number of edges between all nodes with a connected path. We use the Relative

Error (RE) of path to evaluate the error.

$$\mathrm{RE}_P = \frac{|P - \tilde{P}|}{\max(\varphi, P)}, \tag{27}$$

where P and $\tilde{P}$ are the paths of the original and synthetic weighted graphs, respectively, φ is a factor to bound the impact of small results.

5.2 Comparison under w-Event Privacy

For each dataset, we first evaluate the utility of WGT-LDP with two alternative methods BDG and RPG under the privacy guarantee of w-event edge weight LDP. We illustrate the effectiveness of WGT-LDP from two perspectives: different privacy budgets and different sliding windows. Moreover, we show the impact of privacy budget allocation and data range estimation in WGT-LDP.

Utility on Different Privacy Budgets. Figure 3 shows the comparison of four metrics between WGT-LDP and all baselines when ε varies from 0.5 to 2.5. In general, we observe that WGT-LDP outperforms its competitors in most cases. This is because WGT-LDP determines the sampling nodes based on data change and the number of remaining nodes at each time step, which brings two benefits: 1) each node can perturb their data with sufficient privacy budget, reducing the perturbation error; 2) nodes with large change errors are preferentially sampled, reducing the approximation error. Therefore, WGT-LDP preserves the information of the original weighted graph snapshot more accurately than the baseline method. When the privacy budget ε is less than 1.5, WGT-LDP has higher RE on Email-Eu than the baseline RPG. This is because small graphs are more sensitive to noise and too small privacy budget reduces the sampling accuracy. We notice that the KL divergence of the weight distribution of RPG on the Tech-AS is close to WGT-LDP, which may be due to the fact that the weights of most nodes on the Tech-AS vary greatly, resulting in similar approximation errors. For other baselines, we find that they have different relative performances on different metrics and datasets, but their errors are still high compared to WGT-LDP. In some cases, error slightly increases with larger ε due to the interaction between sampling randomness (e.g., in RPG) and graph structure. As noise decreases, the approximation error introduced by random sampling may dominate, leading to unexpected error patterns.

Utility on Different Window Sizes. Figure 4 shows the utility of WGT-LDP and all baselines on four metrics, with different sliding window size w. In these experiments, the privacy budget ε is fixed to 2. We can see that the errors of all methods on the four metrics generally increase with w. This is because the increase of w means that less privacy budget or nodes will be allocated to each time step. We can also observe that WGT-LDP acquires smaller error than the competitors on three datasets, which shows the effectiveness of our population division-based sampling and subsequent graph synthesis phases in improving utility. For HMG, we find that as w increases, its errors on many metrics such as clustering coefficient and path condition are increasingly different from our scheme. In particular, HMG adopts a hidden Markov model to predict the interaction probability between current users. However, the increase of w leads to greater noise, which causes this prediction based on historical data to become very inaccurate. Comparing the baselines

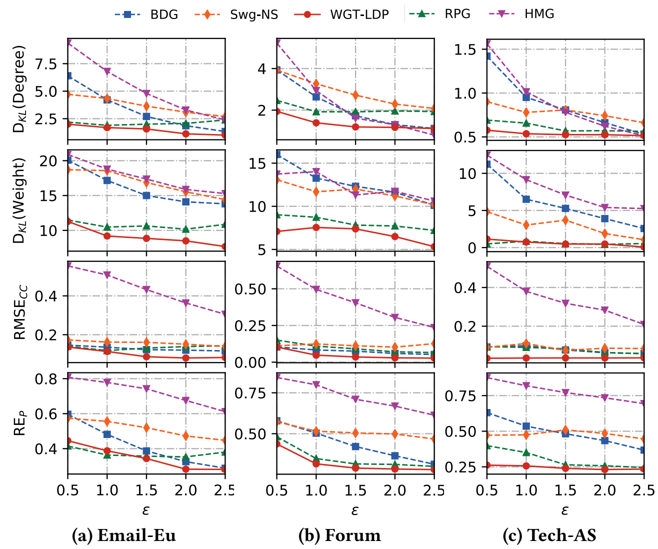


Figure 3: Utility comparison on four metrics given different privacy budgets.

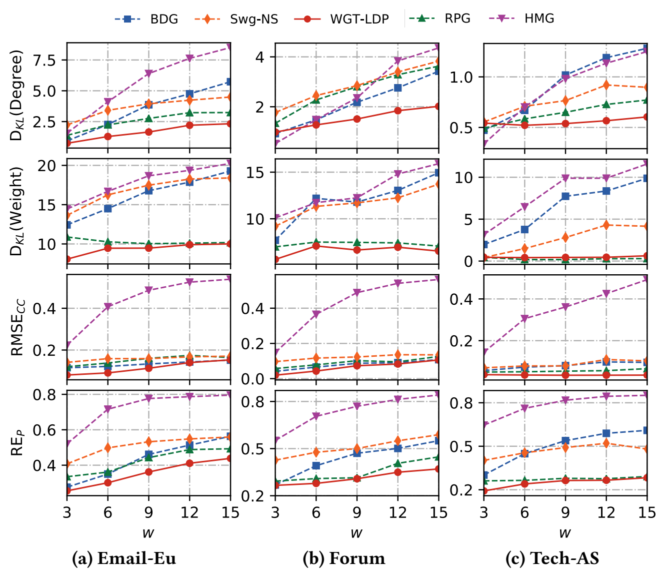


Figure 4: Utility comparison on four metrics given different window sizes.

BDG and RPG, although BDG performs better than RPG on the degree distribution of Forum, it performs worse than RPG on Tech-AS, which suggest that the relative utility of the two baselines depends on the dataset.

Effect of Threshold δ. Recalling Section 3.2, we extract nodes with large data change errors based on the threshold δ at each time step, which determine the number of subsequent sampling nodes. In Figure 5, we show the performance of WGT-LDP under the threshold δ ranging from 0.5 to 9, after fixing the total privacy budget $\varepsilon = 2$. We find that the effect of varying δ on utility tends to

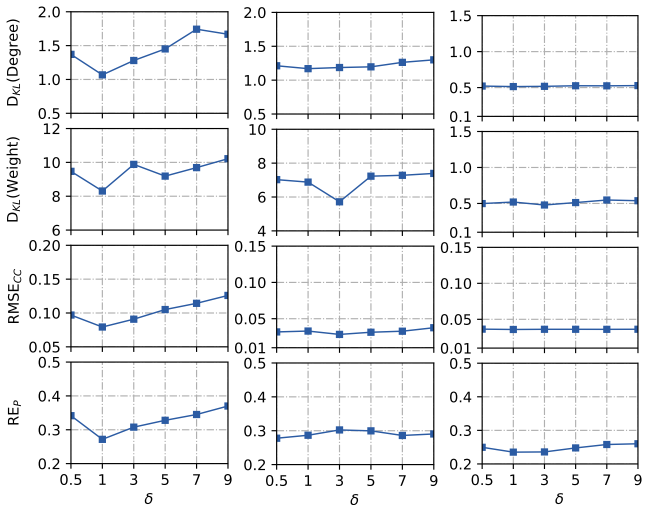


Figure 5: Four evaluation metrics vs. Different threshold δ

level off as the dataset grows. The reason is as follows: When the total number of nodes n increases, the number of nodes m_t with data changes greater than δ increases significantly. At this time, we limit the number of sampling nodes to near $\frac{n}{w}$ by setting the sampling ratio, so that the adjustment of δ has little effect on the determination of the final sampling nodes. For the Email-Eu and Forum datasets, we can observe that as δ increases, many metrics first perform better and then worse, such as degree distribution, weight distribution and clustering coefficient. This is caused by the combined effect of perturbation noise and approximation error. Specifically, when δ is too small, many nodes that are basically unchanged will be sampled, which leads to a larger data dimension, and thus increases noise injection. As δ increases, nodes that really have changed can be screened out to balance perturbation noise and approximation error. However, when δ continues to increase, some nodes that actually have changed are omitted, resulting in a larger approximation error. Therefore, a suitable threshold δ is crucial to balance the impact of the above two aspects. WGT-LDP shows consistently excellent performance when the threshold δ is around 1.

Effect of Privacy Budget Allocation. We evaluate the effect of different privacy budget allocation strategies on three real-world datasets. We vary the ratio of ε_1, ε_2 and ε_3 to ε from 0.1 to 0.7 with step size 0.2. Then ε_4 is automatically calculated by $\varepsilon - \varepsilon_1 - \varepsilon_2 - \varepsilon_3$. Figure 6 shows the performance of WGT-LDP when the total privacy budget ε is 1. We observe that the allocation ratio has a significant impact on the evaluation metrics. When ε_1 is small, the errors of most metrics on all datasets are large. The reason is that the first phase needs to divide ε_1 into w time steps, and a small ε_1 further reduces the sampling accuracy. If the sampled nodes are inaccurate, it is difficult to reconstruct a high-quality synthetic graph. In addition, for each dataset, small ε_2, ε_3 and ε_4 have poor effects on different metrics. The above observations provide guidance for privacy budget allocation. That is, we should allocate more privacy budget to ε_1. In our experiments, we set ε_1 as $\frac{1}{2}\varepsilon$, and set ε_2, ε_3 and ε_4 as $\frac{1}{6}\varepsilon$.

Effect of the Number of Experimental Repetitions. Figure 7 illustrates the average utility of multiple repeated experiments. We can observe that when the number of runs is small, the utility of many metrics will vary greatly, e.g., the KL divergence of degree distribution and weight distribution. However, as the number of runs increases, the results of most methods tend to be stable. This is because the random perturbations of the differential privacy mechanism will cause large fluctuations in the results of a single experiment. Therefore, by averaging multiple independent experiments, the performance of the algorithm in the sense of expectation can be more reliably evaluated. In addition, WGT-LDP achieves the best performance in most cases, which proves the effectiveness of our mechanism on synthetic dynamic weighted graphs.

Effect of Data Range Estimation. We conduct experiments of WGT-LDP with and without data range estimation on three datasets to evaluate the effect of data range estimation. As shown in Figure 8, we can observe that data range estimation significantly reduces the KL divergence of weight distribution. The reason is that WGT-LDP with data range estimation avoids pathological worst case of considering edge weight perturbations by estimating the local upper bound for each node, which reduces the noise of edge weights in any snapshot. When the privacy budget is small, the estimation is not accurate enough, thus causing higher KL divergence.

5.3 Comparison under Event Privacy

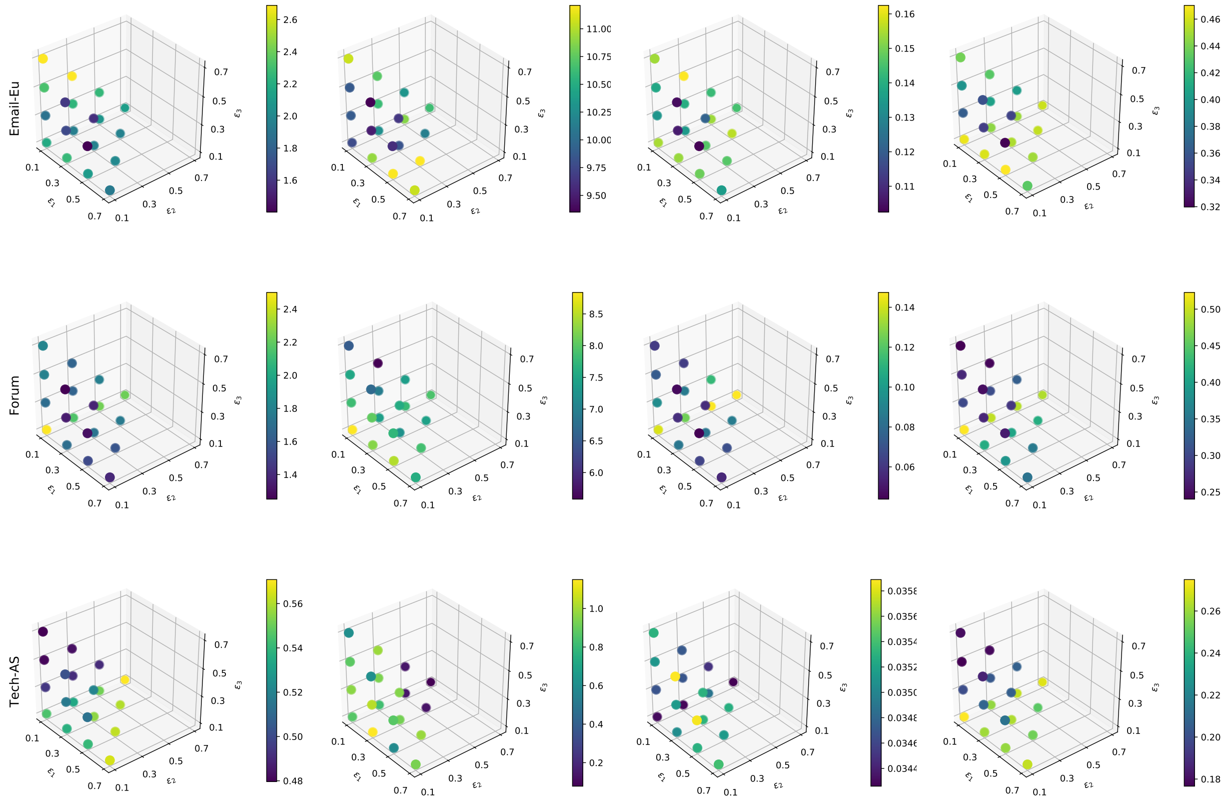
To show the advantages of the last three stages of WGT-LDP for graph generation, we further provide performance comparisons when $w = 1$. In other words, we consider the scenario of event-level edge weight LDP. In this scenario, WGT-LDP protects the privacy of edges and weights on any single graph snapshot, which can be regarded as a static graph publication at any time step. Since static graph mechanisms LDPGen and Blink only focus on unweighted graphs, we compare all methods on three metrics, i.e., degree distribution, path condition, and clustering coefficient.

Figure 9 illustrates the experimental results on three datasets. Since WGT-LDP is designed for generating weighted graphs, it is necessary to allocate additional privacy budget for the reconstruction of edge weights. Nevertheless, we observe that WGT-LDP still achieves better accuracy than the baselines based on unweighted graphs in most cases. The reason is that WGT-LDP exploits the relationship between edge weights and graph structure, which can be used to guide the synthesis of weighted graphs. For clustering coefficient, LDPGen performs best on Forum and Tech-AS because it uses the BTER [39] model to generate a synthetic graph, which is specifically optimized for returning accurate clustering coefficients.

5.4 Evaluation on Synthetic Datasets

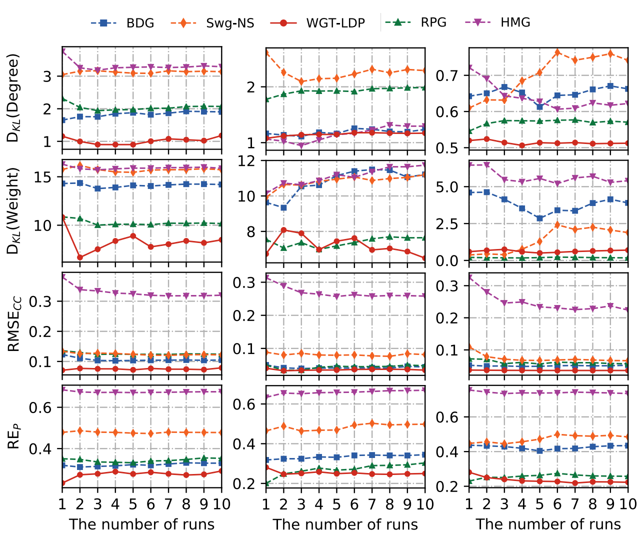
To evaluate the impact of the severity of snapshot changes and the larger window size w on algorithm performance, we compare the utility of WGT-LDP with alternative methods on two synthetic datasets given different privacy budgets and larger w. For efficiency reasons, we omit the baseline BDG due to its high computational overhead at this scale.

Figure 10 shows the performance of various methods on four



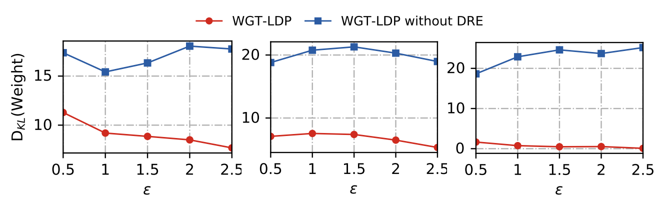
(a) D_{KL}(Degree) (b) D_{KL}(Weight) (c) $RMSE_{CC}$ (d) RE_P

Figure 6: Four evaluation metrics vs. Different privacy budget allocations.



(a) Email-Eu (b) Forum (c) Tech-AS

Figure 7: Average utility of multiple repeated experiments. The number of repetitions ranging from 1 to 10.



(a) Email-Eu (b) Forum (c) Tech-AS

Figure 8: KL divergence of weight distribution with and without data range estimation.

metrics, with different privacy budgets. We can see that in terms of the four metrics, WGT-LDP performs better than other methods on Synthetic-II, but its advantage on Synthetic-I is not obvious. This is because the severity of snapshot changes in Synthetic-II is higher than that in Synthetic-I. For WGT-LDP, when the degree of data change is large, it can effectively capture nodes with large change errors, which improves the utility of graph synthesis. For the baseline RPG, we observed that when the privacy budget is large, the baseline RPG shows an upward trend on some metrics as the privacy budget increases. The reason is that although the large privacy budget limits the impact of injected noise, the random sampling method of RPG causes some areas with important graph

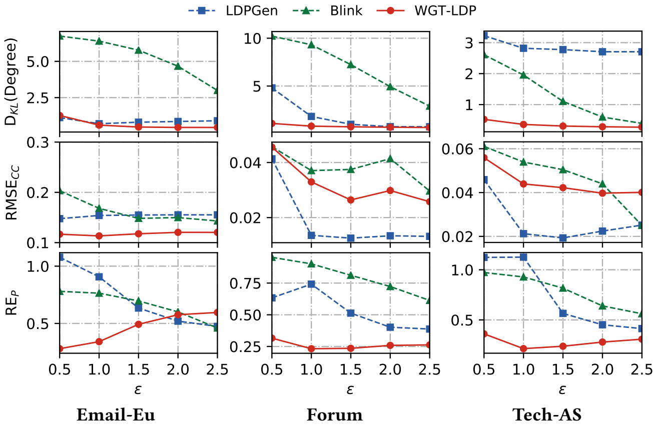


Figure 9: Utility comparison under event-level privacy (i.e., $w = 1$).

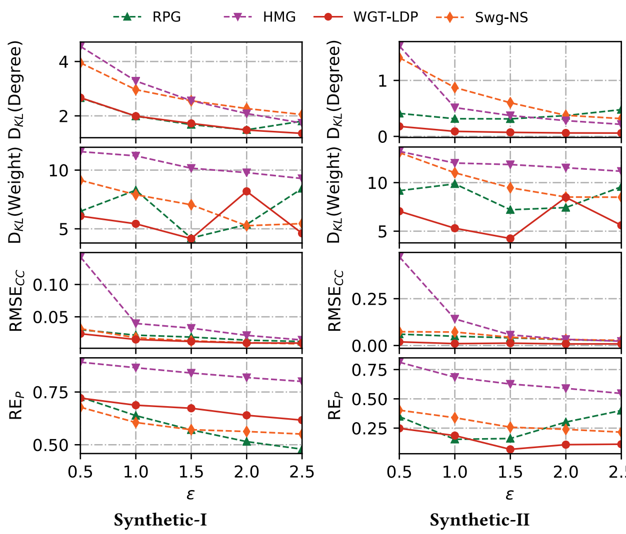


Figure 10: Evaluation on synthetic datasets given different privacy budgets.

features to produce larger approximation errors.

Figure 11 shows the performance of various methods on four metrics when window size w varies from 10 to 90. We also find that WGT-LDP has a more obvious performance advantage on Synthetic-II than on Synthetic-I. In terms of clustering coefficient, the baselines RPG and Swg-NS slightly outperform our method. This is because the clustering coefficient of the synthetic dataset is very low, i.e., there is no significant community structure, and the sampling strategy of the baselines RPG and Swg-NS increases the dispersion of the graph structure, which in turn leads to a smaller RMSE. Nevertheless, the overall performance of WGT-LDP remains competitive, especially on Synthetic-II. In addition, we observe that as w increases, our method first performs better and then worse on weight distribution, which is caused by the combined effect of perturbation error and approximation error. When w is small, more nodes are sampled and we set a larger upper bound on weights in

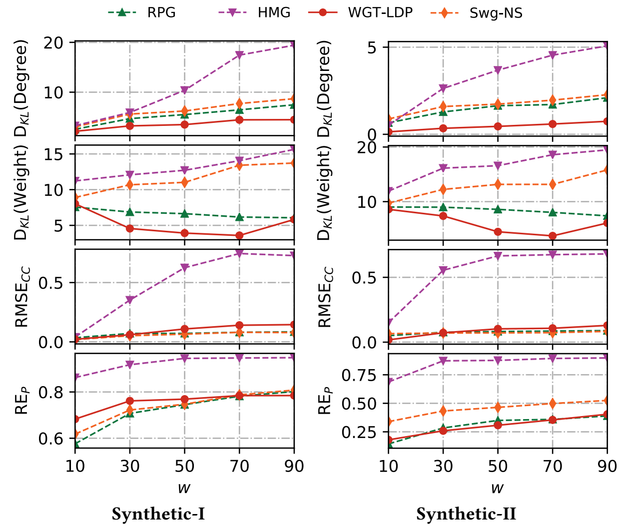


Figure 11: Evaluation on synthetic datasets given different window sizes (from 10 to 90).

the synthetic graph, making the perturbation noise of weights large. When w is large, more data is approximated, which aggravates the approximation error.

5.5 Case Study

We apply WGT-LDP to an end-to-end use case, i.e., the influence maximization (IM) problem, to evaluate the utility of synthetic weighted graph snapshots for network analytics. IM aims to locate nodes from the network that can achieve the maximum impact spread, and has applications in viral marketing [6], epidemic control [5], network monitoring [32], and so on. For example, in the viral marketing scenario, companies try to achieve efficient dissemination of new products or recruitment information by recommending these contents to certain influential individuals. In the event of an infectious disease outbreak (such as COVID-19), super spreaders can be identified to prioritize quarantine or vaccination.

Specifically, we find the top 20 most influential nodes on each synthetic graph snapshot using the classical Degree Discount (DD) method [4], where in weighted graphs, the degree of a node is calculated as the sum of the weights of all its adjacent edges. Then, we adopt the standard Independent Cascade (IC) [24] model to compare the average influence spreads of the nodes selected by WGT-LDP and all baselines on all snapshots. In IC model, the spread probability is set to $1-(1-p)^{a(u,v)}$, where $p = 0.01$ and $a(u, v)$ is the weight of the edge (u, v). Higher influence spread values indicate higher utility, i.e., the locations of the most influential nodes are more accurate.

Figure 12 shows the results of influence spread on three real-world datasets, with different privacy budgets. We can observe that WGT-LDP achieves the best utility under all privacy budgets compared to other baselines. This demonstrates the effectiveness of our scheme in recovering graph structure and edge weights, which is crucial for locating influential nodes. For other baselines, we find that RPG significantly outperforms the other three baselines

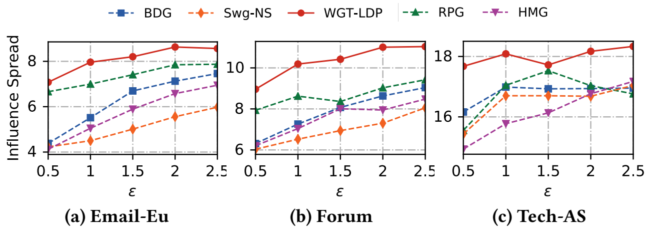


Figure 12: The influence spread given different privacy budgets.

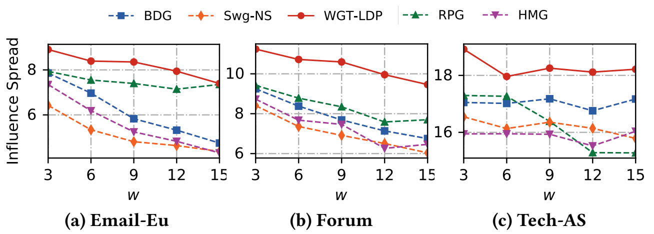


Figure 13: The influence spread given different window sizes.

when the privacy budget is small. This is because BDG, HMG and Swg-NS further split the already small privacy budget, resulting in a significant increase in noise, while the random population division of RPG allows each user to hold the entire privacy budget.

Figure 13 shows the results of influence spread on three real-world datasets, with different sliding window size. WGT-LDP also maintains excellent performance relative to all baselines. In addition, the influence spread of all methods generally decreases with the increase of w, mainly due to the available privacy budget or nodes at each time step decreases, posing a challenge to synthetic graphs.

5.6 Utility on Different Time Steps

To achieve a more comprehensive analysis, we present the utility of WGT-LDP and all baselines at different time steps when the privacy budget is 2. As shown in Figure 14, we observe that WGT-LDP consistently outperforms other methods at most time steps. This highlights the key role played by our node sampling mechanism, weight optimization method, and synthetic graph snapshot step in preserving graph features.

5.7 Running Time

Figure 15 illustrates the average running time on synthetic datasets with node counts $n \in \{1, 5, 10, 15, 20\} \times 10^3$ and maximum edge weight $B = 200$. We set $\varepsilon = 2$ and report the average over 10 time steps. We observe that as the number of nodes increases, the average running time of WGT-LDP exhibits an approximately quadratic growth trend, which is consistent with the dominant term $O(n^2)$ in the computational complexity analysis. Experimental results indicate that the scalability of the proposed method is primarily constrained by $O(n^2)$ under the current experimental setup.

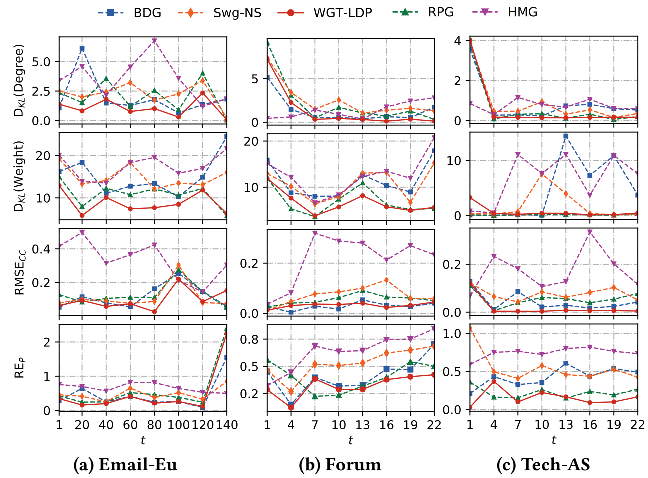


Figure 14: Utility of various methods with different time steps.

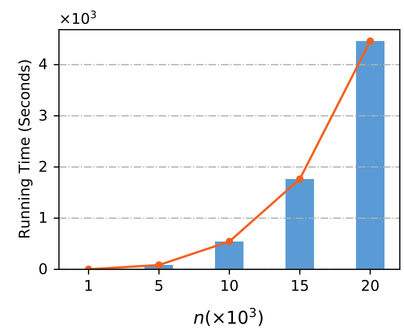


Figure 15: Average running time of WGT-LDP on synthetic datasets with different numbers of nodes.

6 RELATED WORK

Time Series Data Analysis via DP. There are three levels of local differential privacy protection for time series data analysis: event-level, user-level and w-event-level. Event-level LDP aims to hide a single event in a time series [21, 41]. User-level LDP tries to hide all events of a single user [1, 11]. w-event-level LDP aims to protect any event sequence occurring within any window of w time steps. Since w-event-level LDP can balance the privacy loss and utility loss between event-level LDP and user-level LDP, it has been widely used by many works recently [17, 28, 30, 36, 43]. Specifically, Wang et al. [43] proposed a metric-based w-event-level LDP to protect the important patterns of time series. Li et al. [30] extended the above pattern protection to the scenario of finite data range. Ren et al. [36] applied the idea of budget division to the local setting of private streaming data publishing and studied a new population division method for tabular data. Hu et al. [17] explored real-time trajectory synthesis, which generates high-utility trajectory data by extracting movement patterns from users' trajectory streams. However, The algorithms in the above works are tailored to their respective data formats and do not account for graph topology, edge correlations or weight semantics.

Static Graph Publication via DP. Many prior works [2, 3, 19, 20, 35, 44, 45, 48] focus on releasing static graphs under differential privacy guarantees. For instance, Qin et al. [35] first proposed to generate a synthetic social graph under LDP. Wei et al. [44] additionally considered node attributes and generated attributed social graph in a decentralized network. Recently, Yuan et al. [48] utilized community information to publish a synthetic graph with DP, which improves the accuracy of graph reconstruction. Brito et al. [2] designed a DP algorithm to generate count-weighted graph while protecting one interaction between nodes. However, the above studies overlook the temporal properties of graphs, and the proposed methods may be challenging to apply in our continuous publication scenario.

Dynamic Graph Publication via DP. Several works [29, 46, 49] address the temporal or dynamic properties of graphs. Specifically, Hou et al. Yuan et al. [49] proposed a framework for publishing stream graphs under DP, which uses communities as granularity and achieves w-event-level privacy. Xu et al. [46] initially investigated the problem of dynamic weighted graph publication and proposed a new weighted graph snapshot publication mechanism called SwgDP. SwgDP guides current snapshot generation by leveraging historical graph data. However, SwgDP only achieved event-level privacy under DP, which is a rather weak privacy model. That is, individuals can only guarantee the privacy of their edges and weights on any single graph snapshot. Moreover, the DP setting requires the assumption of a trusted curator. Recently, Li et al. [29] considered dynamic graph publishing with LDP for decentralized applications, but their work also only follows the event-level privacy model. In summary, all these studies rely on event-level privacy under DP or LDP, or on w-event privacy in the central DP setting. To the best of our knowledge, there is no prior work on graph data publishing that addresses the combination of dynamic graph structure, edge weights and local w-event privacy, which are necessary to achieve strong protection in dynamic and decentralized scenarios. Our work is the first effort to define and solve the problem of continuous weighted graph publishing with LDP while providing w-event-level privacy.

7 CONCLUSION

In this paper, we propose WGT-LDP, a novel framework for the continuous publication of weighted graph snapshots while ensuring w-event edge weight LDP. By incorporating population division-based sampling, data range estimation, aggregate information collection, and graph snapshot generation, our approach effectively balances privacy protection and utility preservation. Theoretical analysis and extensive experiments on real-world and synthetic datasets demonstrate that WGT-LDP significantly outperforms baseline methods by reducing perturbation noise and improving graph utility.

REFERENCES

[1] Ergute Bao, Yin Yang, Xiaokui Xiao, and Bolin Ding. 2021. CGM: an enhanced mechanism for streaming data collection with local differential privacy. *VLDB* 14, 11 (2021), 2258–2270.

[2] Felipe T Brito, Victor AE Farias, Cheryl Flynn, Subhabrata Majumdar, Javam C Machado, and Divesh Srivastava. 2023. Global and local differentially private release of count-weighted graphs. *SIGMOD* 1, 2 (2023), 1–25.

[3] Rui Chen, Benjamin CM Fung, Philip S Yu, and Bipin C Desai. 2014. Correlated network data publication via differential privacy. *The VLDB Journal* 23 (2014), 653–676.

[4] Wei Chen, Yajun Wang, and Siyu Yang. 2009. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 199–208.

[5] Reuven Cohen, Shlomo Havlin, and Daniel Ben-Avraham. 2003. Efficient immunization strategies for computer networks and populations. *Physical review letters* 91, 24 (2003), 247901.

[6] Pedro Domingos and Matt Richardson. 2001. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. 57–66.

[7] John C Duchi, Michael I Jordan, and Martin J Wainwright. 2013. Local privacy and statistical minimax rates. In *2013 IEEE 54th annual symposium on foundations of computer science*. IEEE, 429–438.

[8] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*. Springer, 265–284.

[9] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*. Springer, 265–284.

[10] Ken TD Eames, Jonathan M Read, and W John Edmunds. 2009. Epidemic prediction and control in weighted networks. *Epidemics* 1, 1 (2009), 70–76.

[11] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. 2019. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2468–2479.

[12] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429, 6988 (2004), 180–184.

[13] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. 2009. Universally utility-maximizing privacy mechanisms. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*. 351–360.

[14] Aric Hagberg, Pieter J Swart, and Daniel A Schult. 2008. *Exploring network structure, dynamics, and function using NetworkX*. Technical Report. Los Alamos National Laboratory (LANL), Los Alamos, NM (United States).

[15] Sudheendra Hangal, Diana MacLean, Monica S Lam, and Jeffrey Heer. 2010. All friends are not equal: Using weights in social graphs to improve search. In *Workshop on Social Network Mining & Analysis, ACM KDD*, Vol. 130.

[16] Michael Hay, Chao Li, Gerome Miklau, and David Jensen. 2009. Accurate estimation of the degree distribution of private networks. In *2009 Ninth IEEE International Conference on Data Mining*. IEEE, 169–178.

[17] Yujia Hu, Yuntao Du, Zhikun Zhang, Ziquan Fang, Lu Chen, Kai Zheng, and Yunjun Gao. 2024. Real-Time Trajectory Synthesis with Local Differential Privacy. *arXiv preprint arXiv:2404.11450* (2024).

[18] Jacob Imola, Takao Murakami, and Kamalika Chaudhuri. 2021. Locally differentially private analysis of graph statistics. In *USENIX Security*. 983–1000.

[19] Xun Jian, Yue Wang, and Lei Chen. 2021. Publishing graphs under node differential privacy. *IEEE Transactions on Knowledge and Data Engineering* 35, 4 (2021), 4164–4177.

[20] Zach Jorgensen, Ting Yu, and Graham Cormode. 2016. Publishing attributed social graphs with formal privacy guarantees. In *SIGMOD*. 107–122.

[21] Matthew Joseph, Aaron Roth, Jonathan Ullman, and Bo Waggoner. 2018. Local differential privacy for evolving data. *NeuIPS* 31 (2018).

[22] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What can we learn privately? *SIAM J. Comput.* 40, 3 (2011), 793–826.

[23] Georgios Kellaris, Stavros Papadopoulos, Xiaokui Xiao, and Dimitris Papadias. 2014. Differentially private event sequences over infinite streams. (2014).

[24] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. 137–146.

[25] Solomon Kullback. 1997. *Information theory and statistics*. Courier Corporation.

[26] Youhuan Li, Lei Zou, M Tamer Özsu, and Dongyan Zhao. 2020. Space-Efficient Subgraph Search Over Streaming Graph With Timing Order Constraint. *TKDE* 34, 9 (2020), 4453–4467.

[27] Zitao Li, Tianhao Wang, Milan Lopuhaä-Zwakenberg, Ninghui Li, and Boris Škoric. 2020. Estimating numerical distributions under local differential privacy. In *SIGMOD*. 621–635.

[28] Zhetao Li, Junru Wu, Saiqin Long, Zhirun Zheng, Chengxin Li, and Mianxiong Dong. 2024. User-Driven Privacy-Preserving Data Streams Release for Multi-Task Assignment in Mobile Crowdsensing. *TMC* (2024).

[29] Zhetao Li, Yong Xiao, Haolin Liu, Xiaofei Liao, Ye Yuan, and Junzhao Du. 2025. Dynamic Graph Publication with Differential Privacy Guarantees for Decentralized Applications. *IEEE Trans. Comput.* (2025).

[30] Zhetao Li, Xiyu Zeng, Yong Xiao, Chengxin Li, Wentai Wu, and Haolin Liu. 2024. Pattern-sensitive Local Differential Privacy for Finite-Range Time-series Data in Mobile Crowdsensing. *TMC* (2024).

[31] Frank D McSherry. 2009. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *SIGMOD*. 19–30.

[32] Adilson E Motter. 2004. Cascade control and defense in complex networks. *Physical Review Letters* 93, 9 (2004), 098701.

[33] Tore Opsahl. 2013. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social networks* 35, 2 (2013), 159–167.

[34] Ashwin Paranjape, Austin R Benson, and Jure Leskovec. 2017. Motifs in temporal networks. In *WSDM*. 601–610.

[35] Zhan Qin, Ting Yu, Yin Yang, Issa Khalil, Xiaokui Xiao, and Kui Ren. 2017. Generating synthetic decentralized social graphs with local differential privacy. In *CCS*. 425–438.

[36] Xuebin Ren, Liang Shi, Weiren Yu, Shusen Yang, Cong Zhao, and Zongben Xu. 2022. LDP-IDS: Local differential privacy for infinite data streams. In *SIGMOD*. 1064–1077.

[37] Ryan Rossi and Nesreen Ahmed. 2015. The network data repository with interactive graph analytics and visualization. In *AAAI*, Vol. 29.

[38] Tara Seals. 2017. Data breaches increase 40% in 2016. https://www.infosecurity-magazine.com/news/data-breaches-increase-40-in-2016/.

[39] Comandur Seshadhri, Tamara G Kolda, and Ali Pinar. 2012. Community structure and scale-free collections of Erdős-Rényi graphs. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 85, 5 (2012), 056109.

[40] Connor Wagaman, Palak Jain, and Adam Smith. 2024. Time-Aware Projections: Truly Node-Private Graph Statistics under Continual Observation. In *2024 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, 237–237.

[41] Tianhao Wang, Joann Qiongna Chen, Zhikun Zhang, Dong Su, Yueqiang Cheng, Zhou Li, Ninghui Li, and Somesh Jha. 2021. Continuous release of data streams under both centralized and local differential privacy. In *CCS*. 1237–1253.

[42] Tianhao Wang, Milan Lopuhaä-Zwakenberg, Zitao Li, Boris Skoric, and Ninghui Li. 2019. Locally differentially private frequency estimation with consistency. *arXiv preprint arXiv:1905.08320* (2019).

[43] Zhibo Wang, Wenxin Liu, Xiaoyi Pang, Ju Ren, Zhe Liu, and Yongle Chen. 2020. Towards pattern-aware privacy-preserving real-time data collection. In *INFOCOM*. IEEE, 109–118.

[44] Chengkun Wei, Shouling Ji, Changchang Liu, Wenzhi Chen, and Ting Wang. 2020. AsgLDP: Collecting and generating decentralized attributed graphs with local differential privacy. *TIFS* 15 (2020), 3239–3254.

[45] Qian Xiao, Rui Chen, and Kian-Lee Tan. 2014. Differentially private network data release via structural inference. In *SIGKDD*. 911–920.

[46] Wen Xu, Zhetao Li, Haolin Liu, Yunjun Gao, Xiaofei Liao, and Kenli Li. 2024. Differentially Private Weighted Graphs Publication Under Continuous Monitoring. *TMC* (2024).

[47] Qingqing Ye, Haibo Hu, Man Ho Au, Xiaofeng Meng, and Xiaokui Xiao. 2022. LF-GDPR: A Framework for Estimating Graph Metrics With Local Differential Privacy. *IEEE Transactions on Knowledge and Data Engineering* 34, 10 (2022), 4905–4920.

[48] Quan Yuan, Zhikun Zhang, Linkang Du, Min Chen, Peng Cheng, and Mingyang Sun. 2023. PrivGraph: Differentially Private Graph Data Publication by Exploiting Community Information. In *USENIX Security*. 3241–3258.

[49] Quan Yuan, Zhikun Zhang, Linkang Du, Min Chen, Mingyang Sun, Yunjun Gao, Michael Backes, Shibo He, and Jiming Chen. 2024. PSGraph: Differentially Private Streaming Graph Synthesis by Considering Temporal Dynamics. *arXiv preprint arXiv:2412.11369* (2024).

[50] Xiaochen Zhu, Vincent YF Tan, and Xiaokui Xiao. 2023. Blink: Link Local Differential Privacy in Graph Neural Networks via Bayesian Estimation. In *CCS*. 2651–2664.