# Novelty Assessment Report

**Paper**: Differentially Private Domain Discovery
**PDF URL**: https://openreview.net/pdf?id=yBpzF8hp3J
**Venue**: ICLR 2026 Conference Submission
**Year**: 2026
**Report Generated**: 2026-01-05

## Abstract

We study several problems in differentially private domain discovery, where each user holds a subset of items from a shared but unknown domain, and the goal is to output an informative subset of items. For set union, we show that the simple baseline Weighted Gaussian Mechanism (WGM) has a near-optimal $\ell_1$ missing mass guarantee on Zipfian data as well as a distribution-free $\ell_\infty$ missing mass guarantee. We then apply the WGM as a domain-discovery precursor for existing known-domain algorithms for private top-K and K-hitting set and obtain new utility guarantees for their unknown domain variants. Finally, experiments demonstrate that all of our WGM-based methods are competitive with or outperform existing baselines for all three problems.



## Core Task Landscape

This paper addresses: **Differentially Private Domain Discovery with Unknown Item Domains**

A total of **8 papers** were analyzed and organized into a taxonomy with **9 categories**.

### Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Core Domain Discovery and Set Union Mechanisms**
- **Top-k Selection Over Unknown Domains**
- **Histogram Release and Continual Observation**
- **Local Differential Privacy for Frequency Estimation**
- **Bounded Noise Mechanisms for Valid Outputs**

### Complete Taxonomy Tree

- Differentially Private Domain Discovery with Unknown Item Domains Survey Taxonomy
- Core Domain Discovery and Set Union Mechanisms
  - Weighted Gaussian Mechanisms for Set Union ★ (1 papers)
  - [0] Differentially Private Domain Discovery (Anon et al., 2026) View paper
  - Adaptive Partition Selection (1 papers)
  - [7] Scalable Private Partition Selection via Adaptive Weighting (Chen, 2025) View paper
- Top-k Selection Over Unknown Domains
  - Stability-Based Top-k Selection (1 papers)
  - [2] Differentially private top-k selection via stability on unknown domain (Ricardo Silva-Carvalho, 2020) View paper
  - Pay-What-You-Get Composition for Top-k (1 papers)
  - [3] Practical differentially private top-k selection with pay-what-you-get composition (Durfee, 2019) View paper
- Histogram Release and Continual Observation
  - Streaming Histogram Release (1 papers)
  - [5] Differentially Private Histograms under Continual Observation: Streaming Selection into the Unknown (Cardoso, 2021) View paper
  - Unifying Analysis Frameworks for Unknown Domains (1 papers)
  - [4] A Unifying Privacy Analysis Framework for Unknown Domain Algorithms in Differential Privacy (Rogers, 2023) View paper
- Local Differential Privacy for Frequency Estimation
  - Target Range Frequency Estimation (1 papers)
  - [6] When Focus Enhances Utility: Target Range LDP Frequency Estimation and Unknown Item Discovery (Jiang Bo, 2024) View paper
  - Two-Party Heavy Hitters Discovery (1 papers)
  - [1] Differentially Private Two-Party Top- Frequent Item Mining (W Tong, 2023) View paper
- Bounded Noise Mechanisms for Valid Outputs (1 papers)
  - [8] The Bounded Gaussian Mechanism for Differential Privacy (Bo Chen, 2022) View paper

### Narrative

Core task: Differentially private domain discovery with unknown item domains. This field addresses the challenge of privately identifying and releasing information about data domains when the set of possible items is not known in advance. The taxonomy reveals several interconnected research directions. Core Domain Discovery and Set Union Mechanisms focus on fundamental techniques for aggregating sets while preserving privacy, often employing weighted noise injection strategies. Top-k Selection Over Unknown Domains tackles the problem of identifying the most frequent items without prior knowledge of the domain, with works like Practical Private Top-k[3] and Private Top-k Stability[2] exploring stability-based approaches. Histogram Release and Continual Observation addresses dynamic settings where data arrives over time, as seen in Streaming Private Histograms[5]. Local Differential Privacy for Frequency Estimation considers the more stringent local model where users perturb their own data before sharing. Finally, Bounded Noise Mechanisms for Valid Outputs, exemplified by Bounded Gaussian Mechanism[8], ensure that noisy outputs remain within valid ranges.

Several active themes emerge across these branches. A central tension involves balancing utility and privacy when the domain size is unknown or potentially unbounded, requiring adaptive strategies that allocate privacy budget efficiently. Works in top-k selection emphasize stability and practical performance, while histogram release methods must handle continual updates without exhausting privacy budgets. Private Domain Discovery[0] sits within the Core Domain Discovery branch, specifically focusing on weighted Gaussian mechanisms for set union operations. Its emphasis on handling unknown domains through carefully calibrated noise aligns closely with the foundational concerns of Unknown Domain Privacy Framework[4], while its mechanism design shares methodological connections with Bounded Gaussian Mechanism[8]. Compared to top-k selection works like Practical Private Top-k[3], Private Domain Discovery[0] addresses the broader problem of discovering entire domains rather than selecting a fixed number of top items, representing a complementary approach to unknown-domain privacy challenges.

## Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

### Taxonomy-Level Summary

Both subtopics address differentially private domain discovery from unbounded item universes, using noise mechanisms to protect privacy while identifying relevant items. The original leaf focuses on weighted Gaussian noise mechanisms for set union operations with missing mass guarantees, while the sibling uses adaptive weighting strategies for partition selection under user-level privacy constraints. The key distinction lies in adaptivity: one employs fixed weighting schemes for set union, the other dynamically adjusts weights for scalable item selection.

**Similarities:** - Both tackle domain discovery problems where the item universe is unknown or unbounded - Both employ weighting schemes in their privacy mechanisms - Both aim to achieve differential privacy guarantees while discovering relevant items - Both address scalability challenges in private data collection

**Differences:** - Original leaf uses non-adaptive Gaussian mechanisms for set union; sibling uses adaptive weighting for partition selection - Original leaf focuses on missing mass guarantees; sibling emphasizes user-level privacy - Original leaf applies fixed weighting schemes; sibling dynamically adjusts weights based on data - Sibling explicitly targets unbounded universes with adaptive algorithms; original leaf focuses on set union operations

**Suggested Search Directions:** - Hybrid approaches combining adaptive and non-adaptive weighting for domain discovery - Comparative analysis of Gaussian vs. other noise mechanisms (e.g., Laplace) for weighted set union - Extensions to event-level vs. user-level privacy trade-offs in domain discovery - Applications of weighted mechanisms to specific domains (e.g., location data, search queries)

### Sibling Subtopics

- **Adaptive Partition Selection** (leaves: 1, papers: 1)
- Scope: Scalable algorithms using adaptive weighting to select items from unbounded universes for user-level privacy.
- Exclude: Excludes non-adaptive set union methods; see Weighted Gaussian Mechanisms subcategory.

## Contributions Analysis

**Overall novelty summary.** The paper contributes near-optimal missing mass guarantees for the Weighted Gaussian Mechanism (WGM) on Zipfian data and extends WGM to unknown-domain variants of top-k and k-hitting set problems. It sits in the 'Weighted Gaussian Mechanisms for Set Union' leaf under 'Core Domain Discovery and Set Union Mechanisms'. Notably, this leaf contains only the original paper itself—no sibling papers are present. This suggests the specific focus on WGM-based set union with missing mass analysis occupies a relatively sparse position within the broader domain discovery landscape.

The taxonomy reveals neighboring research directions that contextualize this work. The sibling leaf 'Adaptive Partition Selection' addresses scalable item selection from unbounded universes using adaptive weighting, which shares methodological overlap but excludes non-adaptive set union approaches. The adjacent branch 'Top-k Selection Over Unknown Domains' contains stability-based and composition-based methods for ranking frequent items, representing a complementary problem formulation. The 'Histogram Release and Continual Observation' branch tackles dynamic settings with streaming data, while 'Local Differential Privacy for Frequency Estimation' explores client-side randomization without trusted curators—both distinct from the centralized set union focus here.

Among the three contributions analyzed, the literature search examined four candidate papers total. The contribution 'WGM-based algorithms for unknown domain top-k and k-hitting set' was assessed against all four candidates, with zero refutable overlaps identified. The other two contributions—near-optimal missing mass guarantees on Zipfian data and reframing set union utility via missing mass—were not matched against any candidates in the search. This limited scope (four papers examined, not dozens) suggests the analysis captures closely related work but does not constitute an exhaustive survey of all potential prior art in domain discovery or weighted noise mechanisms.

Given the sparse taxonomy position and limited search scope, the work appears to occupy a relatively underexplored niche within differentially private domain discovery. The absence of sibling papers and zero refutable overlaps among examined candidates suggest the specific combination of WGM, missing mass analysis, and unknown-domain extensions may be novel. However, the small candidate pool (four papers) means broader connections to related mechanism design or domain discovery literature may exist beyond this analysis.

---

This paper presents **3 main contributions**, each analyzed against relevant prior work:

### Contribution 1: Near-optimal missing mass guarantees for WGM on Zipfian data

**Description**: The authors prove that the Weighted Gaussian Mechanism achieves near-optimal $\ell_1$ missing mass bounds for datasets exhibiting Zipfian (power-law) frequency distributions, and also establish distribution-free $\ell_\infty$ missing mass guarantees. These are the first absolute utility guarantees for differentially private set union.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### Contribution 2: WGM-based algorithms for unknown domain top-k and k-hitting set

**Description**: The authors develop a meta-algorithm that uses WGM for domain discovery followed by known-domain algorithms for top-k selection and k-hitting set problems. They prove utility guarantees for these unknown domain variants by leveraging their $\ell_\infty$ missing mass bounds.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

#### 1. Practical differentially private top-k selection with pay-what-you-get composition

**URL**: View paper

**Brief Assessment**

Practical Private Top-k[3] focuses on top-k selection in unknown domains but does not use WGM for domain discovery. Instead, it preprocesses data to access only the top-k elements and applies the exponential mechanism with Gumbel noise, which is a fundamentally different algorithmic approach than the WGM-based meta-algorithm described in the original paper.

---

### 2. Enabling Privacy-Preserving Top-k Hamming Distance Query on the Cloud

**URL**: View paper

**Brief Assessment**

Private Hamming Distance[10] addresses privacy-preserving top-k Hamming distance queries on encrypted cloud data using Paillier encryption. The original paper focuses on differentially private top-k selection and k-hitting set problems in unknown domains using the Weighted Gaussian Mechanism. These are fundamentally different problem settings with different privacy models and technical approaches.

---

### 3. A Joint Exponential Mechanism For Differentially Private Top-k

**URL**: View paper

**Brief Assessment**

Joint Exponential Top-k[9] addresses a different problem setting. It assumes a known domain of d items and focuses on efficiently sampling from a joint exponential mechanism. The original paper's contribution involves using WGM for domain discovery in unknown domain settings, which is not addressed by this candidate.

---

### 4. Differentially Private Histograms under Continual Observation: Streaming Selection into the Unknown

**URL**: View paper

**Brief Assessment**

Streaming Private Histograms[5] focuses on continual observation settings with streaming data and polylog(T) privacy loss scaling. The original paper addresses single-shot unknown domain problems with different utility guarantees (missing mass bounds).

---

## Contribution 3: Reframing set union utility in terms of missing mass

**Description**: The authors introduce a new perspective for evaluating differentially private set union by measuring the fraction of total item mass recovered rather than counting unique items. This reframing enables them to prove meaningful utility guarantees where previous cardinality-based approaches could not.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

# Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

# References


- [0] Differentially Private Domain Discovery View paper
- [1] Differentially Private Two-Party Top- Frequent Item Mining View paper
- [2] Differentially private top-k selection via stability on unknown domain View paper
- [3] Practical differentially private top-k selection with pay-what-you-get composition View paper
- [4] A Unifying Privacy Analysis Framework for Unknown Domain Algorithms in Differential Privacy View paper
- [5] Differentially Private Histograms under Continual Observation: Streaming Selection into the Unknown View paper
- [6] When Focus Enhances Utility: Target Range LDP Frequency Estimation and Unknown Item Discovery View paper
- [7] Scalable Private Partition Selection via Adaptive Weighting View paper
- [8] The Bounded Gaussian Mechanism for Differential Privacy View paper
- [9] A Joint Exponential Mechanism For Differentially Private Top-k View paper
- [10] Enabling Privacy-Preserving Top-k Hamming Distance Query on the Cloud View paper