Novelty Assessment Report

Paper: CheckMate! Watermarking Graph Diffusion Models in Polynomial Time
PDF URL: https://openreview.net/pdf?id=92fliNrbxY
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-05

Abstract

Watermarking provides an effective means for data governance. However, conventional post-editing graph watermarking approaches degrade the graph quality and involve NP-hard subroutines. Alternatively, recent approaches advocate for embedding watermarking patterns in the noisy latent during data generation from diffusion models, but remain uncharted for graph models due to the hardness of inverting the graph diffusion process. In this work, we propose CheckWate: the first watermarking framework for graph diffusion models embedding checkerboard watermark and providing polynomial time verification. To address NP-completeness due to graph isomorphism, CheckWate embeds the watermark into the latent eigenvalues, which are isomorphism-invariant. To detect the watermark through reversing the graph diffusion process, CheckWate leverages the graph eigenvectors to approximately dequantizes the discrete graph back to the continuous latent, with theoretical guarantees on the detectability and dequantization error. We further introduce a latent sparsification mechanism to enhance the robustness of CheckWate against graph modifications. We evaluate CheckWate on four datasets and four graph modification attacks, against three generation time watermark schemes. CheckWate achieves remarkable generation quality while being detectable under strong attacks such as isomorphism, whereas the baselines are unable to detect the watermark. Code available at: https://anonymous.4open.science/r/checkwate.

Core Task Landscape

This paper addresses: **Watermarking Graph Diffusion Models with Polynomial Time Verification**

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:

- **Graph Diffusion Model Watermarking**
- **Graph Neural Network Watermarking**
- **Static Graph Data Watermarking**

Complete Taxonomy Tree

- Watermarking Graph Diffusion Models with Polynomial Time Verification Survey Taxonomy
- Graph Diffusion Model Watermarking
  - Latent Space Eigenvalue Watermarking ★ (1 papers)
  - [0] CheckMate! Watermarking Graph Diffusion Models in Polynomial Time (Anon et al., 2026) View paper
  - Molecular Graph Diffusion Watermarking (1 papers)
  - [8] GUISE: Graph GaUssIan Shading watErmark (Yang Ren-yi, 2024) View paper
  - Zero-Watermarking for Diffusion Models (1 papers)
  - [6] PlugMark: A Plug-in Zero-Watermarking Framework for Diffusion Models (P Chen, 2025) View paper
- Graph Neural Network Watermarking
  - Task-Specific GNN Watermarking (1 papers)
  - [1] GENIE: Watermarking Graph Neural Networks for Link Prediction (Gangwal, 2024) View paper
  - Transferable Self-Supervised GNN Watermarking (1 papers)
  - [2] Transferable Watermarking to Self-supervised Pre-trained Graph Encoders by Trigger Embeddings (Zhao, 2024) View paper
  - Backdoor-Free Multi-Bit GNN Watermarking (1 papers)
  - [7] WGLE:Backdoor-free and Multi-bit Black-box Watermarking for Graph Neural Networks (Li Tingzhi, 2025) View paper
  - Imperceptible Owner-Unique GNN Watermarking (1 papers)
  - [3] An Imperceptible and Owner-unique Watermarking Method for Graph Neural Networks (Linji Zhang, 2024) View paper
- Static Graph Data Watermarking
  - Topology-Preserving Graph Watermarking (1 papers)
  - [4] Graph Watermarks (Zhao Xiaohan, 2022) View paper
  - Dynamic Knowledge Graph Diffusion Watermarking (1 papers)
  - [5] KGMark: A Diffusion Watermark for Knowledge Graphs (Peng Hong-rui, 2025) View paper

Narrative

Core task: watermarking graph diffusion models with polynomial time verification. The field addresses intellectual property protection for graph-structured data and models, organized into three main branches. Graph Diffusion Model Watermarking focuses on embedding verifiable signatures into generative models that produce graphs, often leveraging latent space properties or eigenvalue manipulations to ensure efficient verification. Graph Neural Network Watermarking targets discriminative architectures, embedding triggers or backdoors that can be detected through model behavior on specific inputs. Static Graph Data Watermarking embeds marks directly into fixed graph datasets, modifying topology or node features while preserving utility. These branches reflect a progression from protecting static artifacts to safeguarding dynamic generative processes, with verification complexity emerging as a central concern across all settings.

Recent work reveals contrasting priorities: some methods prioritize imperceptibility and robustness against adversarial removal (e.g., Imperceptible Watermarking[3], GUISE[8]), while others emphasize transferability across model architectures (Transferable Watermarking[2]) or plug-and-play deployment (PlugMark[6]). Knowledge graph and embedding-specific approaches like KGMark[5] and WGLE[7] address domain-specific constraints. CheckMate[0] sits within the latent space eigenvalue watermarking cluster, sharing conceptual ground with GENIE[1] in targeting generative diffusion models but distinguishing itself through polynomial-time verification guarantees—a feature that addresses scalability bottlenecks present in earlier graph watermarking schemes like Graph Watermarks[4]. This emphasis on computational efficiency positions CheckMate[0] as bridging theoretical rigor with practical deployment needs, contrasting with methods that achieve strong empirical robustness but lack formal verification bounds.

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

All three subtopics address watermarking for diffusion models, but target different aspects of the problem. The original leaf focuses on graph diffusion models with eigenvalue-based watermarks offering polynomial-time verification guarantees. Siblings diverge by specializing in molecular 3D graphs with Gaussian shading techniques, or by adopting zero-watermarking approaches that verify ownership without embedding detectable patterns.

Similarities: - All three address watermarking or ownership verification for diffusion models - All exclude trigger-based and backdoor watermarking methods - All operate in contexts where graph or latent structure is relevant - All aim to provide verifiable ownership or authenticity guarantees

Differences: - Original leaf embeds watermarks in latent eigenvalues with isomorphism-invariance and polynomial-time detection; siblings either embed in molecular 3D latent spaces or use zero-watermarking without embedded patterns - Original leaf emphasizes polynomial-time verification guarantees; zero-watermarking focuses on plug-in frameworks without output modification; molecular watermarking uses Gaussian shading in 3D molecular contexts - Original leaf targets general graph diffusion models; molecular sibling restricts to 3D molecular graphs; zero-watermarking sibling is domain-agnostic but excludes graph-specific methods - Original leaf excludes GNN task-specific watermarking; molecular sibling excludes 2D graphs and non-molecular domains; zero-watermarking excludes embedded watermarks and graph-specific methods

Suggested Search Directions: - Investigate whether eigenvalue-based watermarking can be adapted to molecular 3D graphs or combined with Gaussian shading techniques - Explore hybrid approaches that combine zero-watermarking verification with latent eigenvalue embedding for enhanced robustness - Examine computational complexity trade-offs between polynomial-time eigenvalue detection and zero-watermarking verification frameworks

Sibling Subtopics

- **Molecular Graph Diffusion Watermarking** (leaves: 1, papers: 1)
- Scope: Watermarking techniques for 3D molecular graph generation using Gaussian shading in latent diffusion.
- Exclude: Excludes 2D graph watermarking and non-molecular domains; see static graph watermarking.
- **Zero-Watermarking for Diffusion Models** (leaves: 1, papers: 1)
- Scope: Plug-in frameworks verifying diffusion model ownership without embedding detectable patterns in outputs.
- Exclude: Excludes embedded watermarks and graph-specific methods; see GNN watermarking.

Contributions Analysis

Overall novelty summary. The paper proposes CheckWate, a watermarking framework for graph diffusion models that embeds checkerboard patterns in latent eigenvalues with polynomial-time verification. According to the taxonomy, this work occupies the 'Latent Space Eigenvalue Watermarking' leaf under 'Graph Diffusion Model Watermarking', where it appears as the sole paper in its specific category. The broader 'Graph Diffusion Model Watermarking' branch contains only three papers total, suggesting this represents a relatively sparse and emerging research direction within the larger watermarking landscape.

The taxonomy reveals three main branches: graph diffusion model watermarking, GNN model watermarking, and static graph data watermarking. CheckWate's neighboring leaves include molecular graph diffusion watermarking and zero-watermarking approaches, both addressing generative model protection but through different mechanisms (Gaussian shading vs. plug-in verification). The sibling branches focus on discriminative GNN architectures with trigger-based methods and static dataset protection through topology modification. CheckWate bridges theoretical concerns about graph isomorphism with practical generative model protection, occupying a distinct position that combines diffusion model generation with isomorphism-invariant verification.

Among 27 candidates examined across three contributions, no clearly refuting prior work was identified. The core CheckWate framework examined 7 candidates with 0 refutations, while the dequantization mechanism and sparsification enhancement each examined 10 candidates, also with 0 refutations. This suggests that within the limited search scope, the combination of eigenvalue-based watermarking, approximate dequantization with theoretical guarantees, and latent sparsification for graph diffusion models appears novel. However, the analysis explicitly covers top-K semantic matches and citation expansion, not an exhaustive literature review, meaning undiscovered overlaps may exist beyond these 27 papers.

Given the sparse taxonomy structure and absence of refuting candidates among those examined, CheckWate appears to address an underexplored intersection of graph generation, isomorphism-invariant verification, and polynomial-time detectability. The limited search scope (27 papers) and emerging nature of graph diffusion watermarking suggest caution in claiming definitive novelty, though the specific technical combination—eigenvalue embedding with dequantization guarantees—shows no direct precedent within the analyzed literature.

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

Contribution 1: CheckWate: polynomial-time watermarking framework for graph diffusion models

Description: The authors introduce CheckWate, a novel framework that embeds checkerboard patterns into the eigenvalues of noisy latents during graph generation. By leveraging isomorphism-invariant eigenvalues, the method achieves polynomial-time watermark verification, circumventing the NP-hardness associated with graph isomorphism problems.

This contribution was assessed against **7 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Cryptography in Semantic Watermarks: Undetectability and Deployment Implications

URL: View paper

Brief Assessment

Cryptography Semantic Watermarks[33] focuses on semantic watermarking for image generation using latent diffusion models with cryptographic primitives, not graph diffusion models. The candidate addresses watermark undetectability and deployment for images, while the original paper tackles graph-specific challenges like isomorphism-invariance and polynomial-time verification for graph structures.

2. Robust GNN Watermarking via Implicit Perception of Topological Invariants

URL: View paper

Brief Assessment

Topological Invariants Watermarking[32] focuses on watermarking trained GNN models for node/graph classification tasks, not graph diffusion models. The candidate embeds watermarks by training GNNs to perceive topological invariants (algebraic connectivity), while the original embeds checkerboard patterns in the eigenvalues of noisy latents during graph generation from diffusion models.

3. Watermarking Diffusion Model

URL: View paper

Brief Assessment

Watermarking Diffusion Model[30] focuses on watermarking text-to-image latent diffusion models (LDMs) for intellectual property protection, not graph diffusion models. The candidate addresses image generation tasks with text conditioning, while the original paper specifically tackles graph generation with polynomial-time verification through eigenvalue-based watermarking to circumvent graph isomorphism problems.

4. PlugMark: A Plug-in Zero-Watermarking Framework for Diffusion Models

URL: View paper

Brief Assessment

PlugMark[6] focuses on watermarking image diffusion models (Stable Diffusion) through knowledge extraction and boundary representations, not graph diffusion models. The technical approaches are fundamentally different: CheckWate embeds checkerboard patterns in eigenvalues of graph latents to address graph isomorphism, while PlugMark[6] uses a plug-in classifier to extract knowledge from image diffusion models without modifying them.

5. GUISE: Graph GaUssIan Shading watErmark

URL: View paper

Brief Assessment

GUISE[8] focuses on watermarking molecular graph generation using latent 3D graph diffusion models by adapting Gaussian Shading techniques. It does not address the polynomial-time verification challenge or the checkerboard eigenvalue embedding approach that CheckWate introduces for general graph diffusion models.

6. Watermarking Graph Neural Networks by Random Graphs

URL: View paper

Brief Assessment

Random Graphs Watermarking[31] focuses on watermarking Graph Neural Networks (GNNs) for node classification tasks using Erdős-Rényi random graphs as triggers. CheckWate addresses watermarking for graph diffusion models with eigenvalue-based verification to handle graph isomorphism. These are fundamentally different problem domains and methodologies.

7. Embedding Watermarks in Diffusion Process for Model Intellectual Property Protection

URL: View paper

Brief Assessment

Embedding Watermarks Diffusion[29] focuses on watermarking general diffusion models (images, MNIST, CIFAR-10, CelebA) by embedding watermarks in intermediate diffusion steps, not graph diffusion models. The candidate does not address graph isomorphism, eigenvalue-based watermarking, or polynomial-time verification challenges specific to graphs.

Contribution 2: Approximate dequantization mechanism with theoretical guarantees

Description: The authors develop a method to reverse the quantization step in graph diffusion by using eigenvector properties to reconstruct continuous latents from discrete adjacency matrices. They provide theoretical bounds on reconstruction error and watermark detectability, enabling accurate watermark verification without solving NP-complete graph matching problems.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Graph convolutional networks with eigenpooling

URL: View paper

Brief Assessment

Eigenpooling[24] focuses on graph classification using eigenvector-based pooling operators for hierarchical graph representation learning, not on dequantization mechanisms for reversing quantization in graph diffusion models or watermark verification.

2. Eigenvector fluctuations and limit results for random graphs with infinite rank kernels

URL: View paper

Brief Assessment

Eigenvector Fluctuations[20] focuses on eigenvector perturbation analysis for random graphs with infinite rank kernels, addressing statistical inference problems like graphon estimation and latent position testing. This is fundamentally different from the original paper's graph diffusion watermarking context, which addresses reversing quantization in graph generation models to enable watermark detection.

3. Manifold graph signal restoration using gradient graph Laplacian regularizer

URL: View paper

Brief Assessment

Gradient Graph Laplacian[21] focuses on manifold graph signal restoration using gradient operators and graph embeddings for continuous manifold spaces, not on reversing quantization in graph diffusion models or discrete-to-continuous reconstruction for watermark detection.

4. Spectral augmentations for graph contrastive learning

URL: View paper

Brief Assessment

Spectral Augmentations[23] focuses on contrastive learning augmentations for graph neural networks, not on dequantization mechanisms for graph diffusion models or watermarking.

5. Asymmetry in Spectral Graph Theory: Harmonic Analysis on Directed Networks via Biorthogonal Bases (Adjacency-Operator Formulation)

URL: View paper

Brief Assessment

Asymmetry Spectral Graph[22] focuses on biorthogonal Fourier transforms for directed graphs in harmonic analysis, not on dequantization mechanisms for reversing quantization in graph diffusion models or reconstructing continuous latents from discrete adjacency matrices.

6. Directed graph contrastive learning

URL: View paper

Brief Assessment

Directed Graph Contrastive[28] focuses on contrastive learning for directed graphs using Laplacian perturbation, not on dequantization mechanisms for graph diffusion models or watermarking applications.

7. Blind Deconvolution on Graphs: Exact and Stable Recovery

URL: View paper

Brief Assessment

Blind Deconvolution Graphs[25] addresses blind deconvolution on graphs using eigenvector properties for signal reconstruction, but focuses on source localization in diffusion processes rather than reversing quantization in graph diffusion models. The technical contexts differ fundamentally: the original paper deals with discrete-to-continuous transitions in generative graph diffusion models, while the candidate addresses continuous signal recovery from diffused observations.

8. Sign and Basis Invariant Networks for Spectral Graph Representation Learning

URL: View paper

Brief Assessment

Sign Basis Invariant[19] focuses on neural network architectures for processing eigenvectors with sign and basis invariance in spectral graph representation learning, not on dequantization mechanisms for reversing quantization in graph diffusion models or watermark detection.

9. Robust spectral clustering with rank statistics

URL: View paper

Brief Assessment

Rank Statistics Clustering[26] focuses on spectral clustering methods for latent structure recovery in noisy data matrices using rank statistics, not on dequantization mechanisms for graph diffusion models. The candidate paper addresses a fundamentally different problem domain (robust clustering in general data matrices) compared to the original paper's focus on watermarking graph diffusion models with dequantization for discrete-to-continuous reconstruction.

10. SHEEP, a Signed Hamiltonian Eigenvector Embedding for Proximity

URL: View paper

Brief Assessment

SHEEP[27] focuses on signed network embedding using eigenvector properties for proximity measures in signed graphs, not on dequantization mechanisms for graph diffusion models or discrete-to-continuous reconstruction in generative contexts.

Contribution 3: Latent sparsification mechanism for robustness enhancement

Description: The authors propose a sparsification technique that replaces unlikely entries in the reconstructed noisy latent with zeros, constraining eigenvalue distributions to prevent false positives. This mechanism improves watermark detection robustness, particularly under adversarial graph perturbations such as edge or node modifications.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Enforcing sparsity on latent space for robust and explainable representations

URL: View paper

Brief Assessment

Sparsity Latent Space[9] focuses on enforcing sparsity through spike-and-slab priors in generative models for improved interpretability and noise robustness, not on watermark detection robustness against graph modifications. The technical approaches and application domains are fundamentally different.

2. Efficient and Robust Continual Graph Learning for Graph Classification in Biology

URL: View paper

Brief Assessment

Continual Graph Learning[15] focuses on graph sparsification for memory efficiency in continual learning settings, not on latent sparsification for watermark robustness against adversarial perturbations. The sparsification mechanisms serve fundamentally different purposes in distinct problem domains.

3. Robust graph representation learning via neural sparsification

URL: View paper

Brief Assessment

Neural Sparsification[10] focuses on graph structure sparsification by removing task-irrelevant edges in graph neural networks, not on latent space sparsification for watermark detection in diffusion models. The mechanisms operate on fundamentally different domains and objectives.

4. Anogat-sparse-tl: A hybrid framework combining sparsification and graph attention for anomaly detection in attributed networks using the optimized loss â¦

URL: View paper

Brief Assessment

Anogat Sparse[11] focuses on graph sparsification for anomaly detection in attributed networks, not on latent sparsification for watermark robustness in graph diffusion models. The technical contexts are fundamentally different.

5. Sparse but strong: Crafting adversarially robust graph lottery tickets

URL: View paper

Brief Assessment

Adversarial Graph Lottery[17] focuses on sparsifying graph adjacency matrices and GNN weights for adversarial robustness against structure perturbation attacks, not on latent sparsification for watermark detection in graph diffusion models. The technical domains and objectives are fundamentally different.

6. Adaptive Sparsified Graph Learning Framework for Vessel Behavior Anomalies

URL: View paper

Brief Assessment

Sparsified Graph Learning[16] focuses on maritime vessel trajectory anomaly detection using graph sparsification to remove noisy edges in spatial-temporal vessel graphs. The original paper's latent sparsification mechanism addresses watermark robustness in graph diffusion models by constraining eigenvalue distributions. These are fundamentally different technical contexts and objectives.

7. Graph Spiking Attention Network: Sparsity, Efficiency and Robustness

URL: View paper

Brief Assessment

Graph Spiking Attention[12] focuses on sparsifying graph attention coefficients for robustness against graph edge noises in GNNs, not on latent sparsification mechanisms for watermark detection in graph diffusion models.

8. Dual stream fusion link prediction for sparse graph based on variational graph autoencoder and pairwise learning

URL: View paper

Brief Assessment

Dual Stream Fusion[13] focuses on link prediction in sparse graphs using variational graph autoencoders, not on watermarking or latent sparsification for robustness against adversarial graph perturbations in diffusion models.

9. SPARSE: Semantic Tracking and Path Analysis for Attack Investigation in Real-time

URL: View paper

Brief Assessment

SPARSE[14] focuses on attack investigation in cybersecurity systems using semantic tracking and path analysis, not on watermarking graph diffusion models or latent sparsification for robustness against graph modifications in generative models.

10. Sparse Graph Attention Networks

URL: View paper

Brief Assessment

Sparse Graph Attention[18] focuses on learning sparse attention coefficients for graph neural networks to identify noisy edges in graph structure, not on sparsifying latent representations in diffusion models for watermark robustness.

Appendix: Text Similarity Detection

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

References

- [0] CheckMate! Watermarking Graph Diffusion Models in Polynomial Time View paper
- [1] GENIE: Watermarking Graph Neural Networks for Link Prediction View paper
- [2] Transferable Watermarking to Self-supervised Pre-trained Graph Encoders by Trigger Embeddings View paper
- [3] An Imperceptible and Owner-unique Watermarking Method for Graph Neural Networks View paper
- [4] Graph Watermarks View paper
- [5] KGMark: A Diffusion Watermark for Knowledge Graphs View paper
- [6] PlugMark: A Plug-in Zero-Watermarking Framework for Diffusion Models View paper
- [7] WGLE:Backdoor-free and Multi-bit Black-box Watermarking for Graph Neural Networks View paper
- [8] GUISE: Graph GaUssIan Shading watErmark View paper
- [9] Enforcing sparsity on latent space for robust and explainable representations View paper
- [10] Robust graph representation learning via neural sparsification View paper
- [11] Anogat-sparse-tl: A hybrid framework combining sparsification and graph attention for anomaly detection in attributed networks using the optimized loss â¦ View paper
- [12] Graph Spiking Attention Network: Sparsity, Efficiency and Robustness View paper
- [13] Dual stream fusion link prediction for sparse graph based on variational graph autoencoder and pairwise learning View paper
- [14] SPARSE: Semantic Tracking and Path Analysis for Attack Investigation in Real-time View paper
- [15] Efficient and Robust Continual Graph Learning for Graph Classification in Biology View paper
- [16] Adaptive Sparsified Graph Learning Framework for Vessel Behavior Anomalies View paper
- [17] Sparse but strong: Crafting adversarially robust graph lottery tickets View paper
- [18] Sparse Graph Attention Networks View paper
- [19] Sign and Basis Invariant Networks for Spectral Graph Representation Learning View paper
- [20] Eigenvector fluctuations and limit results for random graphs with infinite rank kernels View paper
- [21] Manifold graph signal restoration using gradient graph Laplacian regularizer View paper
- [22] Asymmetry in Spectral Graph Theory: Harmonic Analysis on Directed Networks via Biorthogonal Bases (Adjacency-Operator Formulation) View paper
- [23] Spectral augmentations for graph contrastive learning View paper
- [24] Graph convolutional networks with eigenpooling View paper
- [25] Blind Deconvolution on Graphs: Exact and Stable Recovery View paper
- [26] Robust spectral clustering with rank statistics View paper

- [27] SHEEP, a Signed Hamiltonian Eigenvector Embedding for Proximity View paper
- [28] Directed graph contrastive learning View paper
- [29] Embedding Watermarks in Diffusion Process for Model Intellectual Property Protection View paper
- [30] Watermarking Diffusion Model View paper
- [31] Watermarking Graph Neural Networks by Random Graphs View paper
- [32] Robust GNN Watermarking via Implicit Perception of Topological Invariants View paper
- [33] Cryptography in Semantic Watermarks: Undetectability and Deployment Implications View paper