Novelty Assessment Report

Paper: Graph Random Features for Scalable Gaussian Processes
PDF URL: https://openreview.net/pdf?id=89SQfLguNn
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

We study the application of graph random features (GRFs) – a recently-introduced stochastic estimator of graph node kernels – to scalable Gaussian processes on discrete input spaces. We prove that (under mild assumptions) Bayesian inference with GRFs enjoys $ \mathcal{O}(N^{3/2})$ time complexity with respect to the number of nodes N, with probabilistic accuracy guarantees. In contrast, exact kernels generally incur $\mathcal{O}(N^{3})$. Wall-clock speedups and memory savings unlock Bayesian optimisation with over 1M graph nodes on a single computer chip, whilst preserving competitive performance.

Core Task Landscape

This paper addresses: **Scalable Gaussian Processes on Discrete Input Spaces Using Graph Random Features**
A total of **7 papers** were analyzed and organized into a taxonomy with **5 categories**.

Taxonomy Overview

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

- **Random Feature Methods for Kernel Approximation**
- **Gaussian Process Applications on Graphs**

Complete Taxonomy Tree

- Scalable Gaussian Processes on Discrete Input Spaces Using Graph Random Features Survey Taxonomy
- Random Feature Methods for Kernel Approximation
  - Graph Random Features for Discrete Spaces ★ (2 papers)
  - [0] Graph Random Features for Scalable Gaussian Processes (Anon et al., 2026) View paper
  - [4] General Graph Random Features (Reid, 2023) View paper
  - Variance Reduction via Optimal Transport Couplings (2 papers)
  - [1] Variance-Reducing Couplings for Random Features (Reid, 2024) View paper
- Gaussian Process Applications on Graphs
  - Online and Incremental GP Learning on Graphs (2 papers)
  - [6] Online Graph-Guided Inference Using Ensemble Gaussian Processes of Egonet Features (Konstantinos D. Polyzos, 2021) View paper
  - [7] Ensemble Gaussian Processes for Online Learning Over Graphs With Adaptivity and Scalability (Konstantinos D. Polyzos, 2021) View paper
  - Conformal Prediction for Graph-Based Uncertainty Quantification (1 papers)
  - [2] Conformalized Gaussian processes for online uncertainty quantification over graphs (Xu Jinwen, 2025) View paper
  - GP-Based SLAM and Spatial Mapping (1 papers)
  - [3] GP-SLAM: laser-based SLAM approach based on regionalized Gaussian process map reconstruction (Bo Li, 2020) View paper

Narrative

Core task: Scalable Gaussian processes on discrete input spaces using graph random features. The field addresses the challenge of applying Gaussian process (GP) models to large-scale problems where inputs are discrete or graph-structured, rather than continuous vectors. The taxonomy reveals two main branches: one focused on Random Feature Methods for Kernel Approximation, which develops techniques to approximate expensive kernel computations through sampling strategies, and another on Gaussian Process Applications on Graphs, which adapts GP machinery to graph-based domains such as robotics and network analysis. Within the first branch, works like Variance Reducing Couplings[1] and Optimal Transport Couplings[5] refine the quality of random feature approximations, while General Graph Random Features[4] and Egonet Features[6] extend these ideas to graph-structured inputs. The second branch includes application-driven studies such as GP-SLAM[3] for simultaneous localization and mapping, and methodological contributions like Conformalized Gaussian Processes[2] and Ensemble Gaussian Processes[7] that enhance uncertainty quantification and scalability.

A particularly active line of work centers on designing random features that respect the combinatorial structure of discrete or graph inputs, balancing approximation fidelity with computational efficiency. Graph Random Features[0] sits squarely within this cluster, building on the foundation laid by General Graph Random Features[4] but emphasizing scalability for discrete spaces through novel sampling schemes. Compared to Variance Reducing Couplings[1], which focuses on variance reduction in continuous settings, Graph Random Features[0] tailors its approach to the unique geometry of graphs. Meanwhile, application-oriented works like GP-SLAM[3] demonstrate the practical payoff of scalable GP methods, though they typically operate in continuous spatial domains rather than purely discrete structures. The original paper thus occupies a niche at the intersection of kernel approximation theory and discrete optimization, addressing a gap where classical random feature methods meet graph-based inference.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. General Graph Random Features

Authors: Reid, Isaac, Isaac Reid, Choromanski, Krzysztof, et al. (13 authors total) | **Year/Venue**: 2023 | **URL**: View paper

Abstract

We propose a novel random walk-based algorithm for unbiased estimation of arbitrary functions of a weighted adjacency matrix, coined universal graph random features (u-GRFs). This includes many of the most popular examples of kernels defined on the nodes of a graph. Our algorithm enjoys subquadratic time complexity with respect to the number of nodes, overcoming the notoriously prohibitive cubic scaling of exact graph kernel evaluation. It can also be trivially distributed across machines, permi...

Relationship Analysis

Both papers belong to the same taxonomy category focusing on graph random features for discrete input spaces using random walk-based kernel approximation techniques. They share the core approach of using random walks to construct sparse, unbiased estimates of graph node kernels with subquadratic complexity. The key difference is that the original paper specifically applies GRFs to scalable Gaussian processes with detailed analysis of Bayesian inference and optimization (achieving O(N^(3/2)) complexity), while the candidate paper introduces the general GRF framework with neural modulation functions for broader kernel learning applications including ODEs, clustering, and mesh regression.

Contributions Analysis

Overall novelty summary. The paper applies graph random features to scalable Gaussian processes on discrete input spaces, contributing theoretical guarantees and demonstrating large-scale Bayesian optimization. Within the taxonomy, it resides in the 'Graph Random Features for Discrete Spaces' leaf under 'Random Feature Methods for Kernel Approximation'. This leaf contains only two papers total, including the original work, indicating a relatively sparse and emerging research direction focused specifically on random feature techniques for graph-structured inputs.

The taxonomy reveals two main branches: random feature approximation methods and direct GP applications on graphs. The original paper's leaf sits alongside 'Variance Reduction via Optimal Transport Couplings', which addresses variance reduction in random features but not specifically for discrete spaces. The sibling branch 'Gaussian Process Applications on Graphs' contains application-driven work (online learning, conformal prediction, SLAM) that uses GPs on graphs without random feature approximation. The paper thus bridges kernel approximation theory with discrete optimization, occupying a niche distinct from both variance reduction techniques and direct graph GP applications.

Among thirty candidates examined, the first contribution (applying graph random features to scalable GPs) shows one refutable candidate out of ten examined, suggesting some prior work exists in this direction. The second contribution (theoretical O(N^(3/2)) complexity analysis) examined ten candidates with none refutable, indicating potential novelty in the complexity guarantees. The third contribution (scalable Bayesian optimization on massive graphs) also examined ten candidates with none refutable, suggesting this application scale may be new. The limited search scope means these findings reflect top-thirty semantic matches rather than exhaustive coverage.

Based on the top-thirty semantic search results and taxonomy structure, the work appears to advance a sparse research direction with modest prior overlap in its core application but potentially novel theoretical and scale contributions. The analysis covers semantically similar papers but does not claim exhaustive field coverage, particularly for work outside the random feature and graph GP intersection.

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

Contribution 1: Application of graph random features to scalable Gaussian processes

Description: The authors apply graph random features, a Monte Carlo estimator based on random walks, to construct sparse estimates of learnable graph node kernels for use as covariance functions in Gaussian processes on discrete input spaces.

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. General Graph Random Features

URL: View paper

Prior Art Analysis

General Graph Random Features[4] demonstrates that graph random features (GRFs) were already proposed and analyzed prior to the original paper's submission. The candidate paper presents the foundational GRF algorithm for unbiased estimation of arbitrary functions of weighted adjacency matrices, including graph node kernels. While the original paper claims to apply GRFs to Gaussian processes with specific complexity guarantees, General Graph Random Features[4] already established the core GRF methodology, its theoretical properties, and demonstrated its application to kernel-based methods. The original paper's contribution is thus an extension of existing GRF work rather than a novel application of a new technique.

Evidence

Evidence 1 - **Rationale**: This pair shows that General Graph Random Features[4] introduced the GRF algorithm for graph node kernels before the original paper. The original paper explicitly cites this work as 'recently-introduced', acknowledging its prior existence. - **Original**: we propose to instead use the recently-intro - duced class of graph random features (grfs) - sparse, unbiased estimates of graph node kernels computed using random walks (choromanski, 2023; reid et al., 2023). - **Candidate**: we propose a novel random walk-based algorithm for unbiased estimation of arbitrary functions of a weighted adjacency matrix, coinedgeneral graph random features (g-grfs). this includes many of the most popular examples of kernels defined on the nodes of a graph.

Evidence 2 - **Rationale**: Both papers describe using GRFs to construct sparse estimates of graph node kernels. General Graph Random Features[4] already established this methodology, demonstrating that the technique of using GRFs for kernel estimation was not first proposed by the original paper. - **Original**: we use grfs to construct sparse estimates of learnable graph node kernels, and use these as covariance functions for gps. - **Candidate**: our central contribution is a simple modification which generalises the algorithm toarbitrary functions of a weighted adjacency matrix, allowing efficient and unbiased approximation a much broader class of graph node kernels.

Evidence 3 - **Rationale**: Both papers describe GRFs as estimators of power series of weighted adjacency matrices. General Graph Random Features[4] formalized this mathematical framework, showing that the core methodology was established prior to the original paper's work. - **Original**: graph random features . in this paper, we propose to instead use the recently-intro - duced class of graph random features (grfs) - sparse, unbiased estimates of graph node kernels computed using random walks (choromanski, 2023; reid et al., 2023). grfs are monte carlo estimators of power series of... - **Candidate**: consider the matrices kα(w) ∈rnxn, whereα = (αk)∞ k=0 and αk ∈r: kα(w) = ∞∑ k=0 αkwk. (2) we assume that the sum above converges for allw under consideration, which can be ensured with a regulariserw →σw, σ ∈r+.

---

2. Ensemble Gaussian Processes for Online Learning Over Graphs With Adaptivity and Scalability

URL: View paper

Brief Assessment

Ensemble Gaussian Processes[7] focuses on semi-supervised learning with ensemble methods and online learning scenarios, not on using graph random features as Monte Carlo estimators for kernel approximation in the manner proposed by the original paper.

3. Random walk kernels and learning curves for Gaussian process regression on random graphs

URL: View paper

Brief Assessment

Random Walk Kernels[23] focuses on random walk kernels for GP regression on graphs but does not use graph random features as Monte Carlo estimators. The candidate uses kernel eigenvalue methods and belief propagation, not the GRF sparse estimation approach.

4. Graph based Gaussian processes on restricted domains

URL: View paper

Brief Assessment

Restricted Domains[18] focuses on graph Laplacian-based GPs for regression on restricted subsets of Euclidean space, not on graph random features as Monte Carlo estimators for scalable inference. The candidate constructs covariance matrices using eigenpairs of graph Laplacians rather than random walk-based features.

5. Optimization on manifolds via graph Gaussian processes

URL: View paper

Brief Assessment

Manifold Optimization[19] focuses on Bayesian optimization on manifolds using graph GPs with UCB acquisition, not on developing graph random features as a scalable GP method. The candidate does not demonstrate prior work on applying graph random features for scalable GP inference.

6. Variance-Reducing Couplings for Random Features

URL: View paper

Brief Assessment

Variance Reducing Couplings[1] focuses on variance reduction techniques for graph random features through optimal transport couplings, not on the application of GRFs to scalable Gaussian processes as a primary contribution. The candidate paper treats scalable GP inference as a downstream application to demonstrate the benefits of their variance reduction methods, rather than proposing GRFs for scalable GPs as a novel contribution.

7. Scalable graph-based semi-supervised learning through sparse bayesian model

URL: View paper

Brief Assessment

Sparse Bayesian Semi-supervised[20] focuses on semi-supervised learning with graph-based sparse priors for classification tasks, not on using graph random features as Monte Carlo estimators for Gaussian process kernels on discrete spaces.

8. Variance-Reducing Couplings for Random Features: Perspectives from Optimal Transport

URL: View paper

Brief Assessment

Optimal Transport Couplings[5] focuses on variance reduction techniques for random features through optimal transport couplings, not on applying graph random features to construct scalable Gaussian processes on discrete input spaces. The candidate paper's contribution is improving the convergence of existing random feature methods, while the original paper's contribution is the application of GRFs to GP inference with specific complexity guarantees.

9. Kernels and learning curves for Gaussian process regression on random graphs

URL: View paper

Brief Assessment

Learning Curves Random Graphs[22] focuses on theoretical analysis of GP regression learning curves on random graphs using random-walk based kernels, not on the application of graph random features as Monte Carlo estimators for scalable GP inference.

10. Taming graph kernels with random features

URL: View paper

Brief Assessment

Taming Graph Kernels[21] focuses on graph kernels for general kernel methods (e.g., kernel-SVM, k-means clustering) rather than Gaussian processes specifically. The candidate does not address GP inference, Bayesian optimization, or posterior computation—core elements of the original contribution.

Contribution 2: Theoretical analysis with O(N^(3/2)) time complexity guarantees

Description: The authors provide theoretical proofs demonstrating that Bayesian inference using graph random features achieves O(N^(3/2)) time complexity compared to O(N^3) for exact methods, with probabilistic guarantees on approximation quality.

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. Sparse spectral Bayesian permanental process with generalized kernel

URL: View paper

Brief Assessment

Sparse Permanental Process[15] focuses on Bayesian inference for permanental processes using Laplace approximation and Fourier features, achieving O(N^(3/2)) complexity through conjugate gradient methods. However, this work addresses point process intensity estimation rather than general Gaussian process inference on graphs, representing a different application domain and problem formulation.

2. Interpretable bayesian tensor network kernel machines with automatic rank and feature selection

URL: View paper

Brief Assessment

Bayesian Tensor Networks[14] focuses on kernel machines with tensor decompositions for automatic rank selection, not on graph-based Gaussian processes with random features. The computational complexity analysis addresses different algorithmic structures (tensor network updates vs. graph random feature inference).

3. Deep kernel learning

URL: View paper

Brief Assessment

Deep Kernel Learning[17] focuses on combining deep architectures with kernel methods for scalability, achieving O(N) training complexity through KISS-GP approximations and structured algebra. This differs fundamentally from the original paper's theoretical analysis of Bayesian inference using graph random features with O(N^(3/2)) guarantees via conjugate gradients on sparse kernel matrices.

4. Online informative path planning of autonomous vehicles using kernel-based bayesian optimization

URL: View paper

Brief Assessment

Informative Path Planning[16] focuses on environmental information gathering for autonomous vehicles using kernel-based Bayesian optimization with approximate logarithmic complexity, not general Bayesian inference with graph random features achieving O(N^(3/2)) complexity.

5. Fast Bayesian inference with batch Bayesian quadrature via kernel recombination

URL: View paper

Brief Assessment

Batch Bayesian Quadrature[9] focuses on numerical integration for Bayesian inference using kernel recombination, not on Gaussian processes with graph random features. The O(N^(3/2)) complexity in the candidate relates to kernel recombination algorithms for quadrature, not to graph-based GP inference as in the original paper.

6. CARE: Confidence-rich autonomous robot exploration using Bayesian kernel inference and optimization

URL: View paper

Brief Assessment

CARE[11] focuses on robot exploration using Bayesian kernel inference for mutual information prediction, not on general Gaussian process scalability with graph random features. The time complexity analysis in CARE[11] pertains to their specific BKIO method for robot exploration tasks, not to Bayesian inference with graph random features as claimed in the original paper.

7. Approximate Bayesian Kernel Machine Regression via Random Fourier Features for Estimating Joint Health Effects of Multiple Exposures

URL: View paper

Brief Assessment

Bayesian Kernel Regression[8] focuses on approximating Bayesian kernel machine regression for environmental health applications using random Fourier features, not graph-based Gaussian processes. The time complexity improvements discussed relate to reducing MCMC computation time for kernel matrix operations in a different domain (environmental epidemiology with multiple exposures), not graph random features or the specific O(N^(3/2)) theoretical guarantees for graph node kernels.

8. When Gaussian process meets big data: A review of scalable GPs

URL: View paper

Brief Assessment

Scalable GPs Review[10] is a survey paper reviewing existing scalable GP methods. It does not present original theoretical complexity analysis for specific methods like graph random features.

9. Accelerated linearized Laplace approximation for Bayesian deep learning

URL: View paper

Brief Assessment

Accelerated Laplace Approximation[12] focuses on accelerating Bayesian inference for neural networks using linearized Laplace approximation and Nyström kernel approximation, not on Gaussian processes with graph random features. The O(N^(3/2)) complexity in the candidate arises from conjugate gradient iterations for solving linear systems in neural network contexts, which is fundamentally different from the graph-based kernel approximation framework in the original paper.

10. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems

URL: View paper

Brief Assessment

Polynomial Chaos Acceleration[13] focuses on dimensionality reduction for Bayesian inference in inverse problems using Karhunen-Loève expansions and polynomial chaos methods, not on graph-based Gaussian processes or random features for kernel approximations.

Contribution 3: Scalable Bayesian optimisation on massive graphs

Description: The authors demonstrate practical scalability by implementing Bayesian optimisation with Thompson sampling on graphs containing over one million nodes using a single GPU, showcasing the effectiveness of their GRF-based approach.

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. Multi-armed bandits, Thomson sampling and unsupervised machine learning in phylogenetic graph search.

URL: View paper

Brief Assessment

Phylogenetic Graph Search[30] focuses on phylogenetic tree search using multi-armed bandits and Thompson sampling, not Bayesian optimisation on general graph structures with Gaussian processes.

2. Distributed Thompson Sampling Under Constrained Communication

URL: View paper

Brief Assessment

Distributed Thompson Sampling[33] focuses on multi-agent distributed optimization with communication constraints between agents, not on scaling single-agent Bayesian optimization to massive graphs using GPU implementation and graph random features.

3. Graph Neural Thompson Sampling

URL: View paper

Brief Assessment

Graph Neural Thompson[25] focuses on online decision-making with Thompson sampling for graph-structured actions in a bandit setting, not Bayesian optimisation with Gaussian processes on graph nodes. The candidate addresses sequential graph selection rather than GP-based optimisation on massive graph topologies.

4. Thompson sampling for stochastic bandits with graph feedback

URL: View paper

Brief Assessment

Graph Feedback Bandits[28] focuses on Thompson sampling for stochastic bandits with graph feedback structures, not Bayesian optimisation with Gaussian processes on large-scale graphs using GPUs.

5. Mercer features for efficient combinatorial Bayesian optimization

URL: View paper

Brief Assessment

Mercer Features[27] focuses on combinatorial Bayesian optimization over discrete structures (sequences, graphs as input representations) using diffusion kernels, not Gaussian processes on graph nodes with Thompson sampling at the scale demonstrated in the original paper (>1M nodes on a single GPU).

6. Parallel and distributed Thompson sampling for large-scale accelerated exploration of chemical space

URL: View paper

Brief Assessment

Parallel Thompson Sampling[26] focuses on parallel Bayesian optimization for chemical space exploration using neural networks, not graph-based Gaussian processes. The candidate addresses molecular screening with batch sizes up to 500, while the original paper demonstrates GP inference on graphs with over 1 million nodes using GRFs.

7. Bayesian Deep Neural Network-empowered Thompson Sampling for Context-aware Task Offloading in Dynamic Fog Computing

URL: View paper

Brief Assessment

Context-aware Task Offloading[32] focuses on task offloading in fog computing networks using Bayesian deep neural networks with Thompson sampling, not on scalable Bayesian optimisation for massive graphs with over 1 million nodes using GRF-based approaches.

8. BONSAI: Structure-exploiting robust Bayesian optimization for networked black-box systems under uncertainty

URL: View paper

Brief Assessment

BONSAI[29] focuses on robust Bayesian optimization for networked systems under uncertainty, not on scalable GP inference or Thompson sampling on massive graphs with millions of nodes.

9. Scalable Optimization for Wind Farm Control using Coordination Graphs

URL: View paper

Brief Assessment

Wind Farm Control[31] applies Bayesian optimization (Thompson sampling) to wind farm control with coordination graphs, not to general graph-based Gaussian processes. The candidate focuses on a specific application domain (wind turbines) rather than demonstrating scalability of Bayesian optimization on massive general graphs with over 1 million nodes.

10. Online network revenue management using thompson sampling

URL: View paper

Brief Assessment

Network Revenue Management[24] focuses on revenue management with inventory constraints using Thompson sampling for pricing decisions, not on scalable Bayesian optimisation for general functions on massive graphs with over one million nodes.

Appendix: Text Similarity Detection

Textual similarity detection checked 30 papers and found 2 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. General Graph Random Features

Detected in: Core Task (sibling), Contribution: contribution_1

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Graph Random Features for Scalable Gaussian Processes View paper
- [1] Variance-Reducing Couplings for Random Features View paper
- [2] Conformalized Gaussian processes for online uncertainty quantification over graphs View paper
- [3] GP-SLAM: laser-based SLAM approach based on regionalized Gaussian process map reconstruction View paper
- [4] General Graph Random Features View paper
- [5] Variance-Reducing Couplings for Random Features: Perspectives from Optimal Transport View paper
- [6] Online Graph-Guided Inference Using Ensemble Gaussian Processes of Egonet Features View paper

- [7] Ensemble Gaussian Processes for Online Learning Over Graphs With Adaptivity and Scalability View paper
- [8] Approximate Bayesian Kernel Machine Regression via Random Fourier Features for Estimating Joint Health Effects of Multiple Exposures View paper
- [9] Fast Bayesian inference with batch Bayesian quadrature via kernel recombination View paper
- [10] When Gaussian process meets big data: A review of scalable GPs View paper
- [11] CARE: Confidence-rich autonomous robot exploration using Bayesian kernel inference and optimization View paper
- [12] Accelerated linearized Laplace approximation for Bayesian deep learning View paper
- [13] Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems View paper
- [14] Interpretable bayesian tensor network kernel machines with automatic rank and feature selection View paper
- [15] Sparse spectral Bayesian permanental process with generalized kernel View paper
- [16] Online informative path planning of autonomous vehicles using kernel-based bayesian optimization View paper
- [17] Deep kernel learning View paper
- [18] Graph based Gaussian processes on restricted domains View paper
- [19] Optimization on manifolds via graph Gaussian processes View paper
- [20] Scalable graph-based semi-supervised learning through sparse bayesian model View paper
- [21] Taming graph kernels with random features View paper
- [22] Kernels and learning curves for Gaussian process regression on random graphs View paper
- [23] Random walk kernels and learning curves for Gaussian process regression on random graphs View paper
- [24] Online network revenue management using thompson sampling View paper
- [25] Graph Neural Thompson Sampling View paper
- [26] Parallel and distributed Thompson sampling for large-scale accelerated exploration of chemical space View paper
- [27] Mercer features for efficient combinatorial Bayesian optimization View paper
- [28] Thompson sampling for stochastic bandits with graph feedback View paper
- [29] BONSAI: Structure-exploiting robust Bayesian optimization for networked black-box systems under uncertainty View paper
- [30] Multi-armed bandits, Thomson sampling and unsupervised machine learning in phylogenetic graph search. View paper
- [31] Scalable Optimization for Wind Farm Control using Coordination Graphs View paper
- [32] Bayesian Deep Neural Network-empowered Thompson Sampling for Context-aware Task Offloading in Dynamic Fog Computing View paper
- [33] Distributed Thompson Sampling Under Constrained Communication View paper