GENERALIST GRAPH ANOMALY DETECTION VIA A SPECTRAL VOCABULARY

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph anomaly detection (GAD) methods are typically trained on individual graphs, limiting their ability to transfer anomaly knowledge across domains. Developing a generalist GAD model therefore requires representations that support transfer to unseen graphs. A fundamental challenge is the lack of a shared representation space across graphs that effectively leverages diverse graph data while generalizing across domains. To address this challenge, we introduce SV-GAD, a generalist GAD framework built on a shared "graph vocabulary". Our key insight is that spectral-energy statistics characterize how node signals vary with respect to graph structure through common measurements, providing a domain-agnostic representation across graphs. SV-GAD learns transferable anomaly patterns in the spectral space and leverages a mixture-of-experts architecture, where graph filters serve as specialized experts for different spectral behaviors. Extensive experiments across diverse GAD benchmarks demonstrate that SV-GAD effectively transfers anomaly knowledge to unseen graphs and consistently outperforms existing baselines in a zero-shot setting.

1 INTRODUCTION

Graph anomaly detection (GAD) aims to identify nodes whose behaviors deviate from graph structure and attributes, with broad applications in e-commerce, citation networks, and social networks (Li et al., 2022; Qiao et al., 2025b). Despite substantial progress, many GAD methods remain dataset-specific and therefore struggle to generalize directly to heterogeneous graph domains. Such generalization is particularly important in practice. Some newly encountered graphs typically lack anomaly labels and may exhibit substantially different structural and attribute distributions from previously observed ones, making graph-specific models struggle to transfer. This limitation motivates generalist graph anomaly detection (GGAD), which aims to transfer anomaly knowledge learned from source graphs to unseen target graphs (Qiao et al., 2025b).

Cross-domain transfer in GGAD is challenging because graph datasets can differ substantially in attribute dimensionality, feature semantics, and graph structure, making it difficult to construct a representation with consistent meaning across domains. Existing methods address this heterogeneity from several perspectives. ARC (Liu et al., 2024b) and UNPrompt (Niu et al., 2024) align heterogeneous node attributes via successive feature transformations to obtain a shared-dimensional representation space (Duan et al., 2024). While these approaches address dimensional mismatch, matching dimensions alone does not ensure semantic alignment, as the same coordinate may encode different information across datasets. IA-GGAD (Zhang et al., 2026) instead learns domain-invariant node representations together with affinity features to reduce both feature-space and structural shifts, whereas ReFi-GAD (Liu et al., 2026b) avoids direct raw-feature alignment by constructing predefined relational fingerprints that characterize contextual and structural relationships between nodes and their neighborhoods. Despite these advances, a common challenge remains: existing representations are either tied to dataset-specific feature transformations, rely on learned universal embeddings, or depend on predefined anomaly statistics and assumptions. Consequently, they do not explicitly provide a shared representation whose semantics are defined consistently with respect to graph structure across heterogeneous domains.

To address this gap, we introduce a spectral perspective for cross-domain GAD. Spectral quantities characterize how graph signals vary with respect to graph structure Tang et al. (2022); Liu et al.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

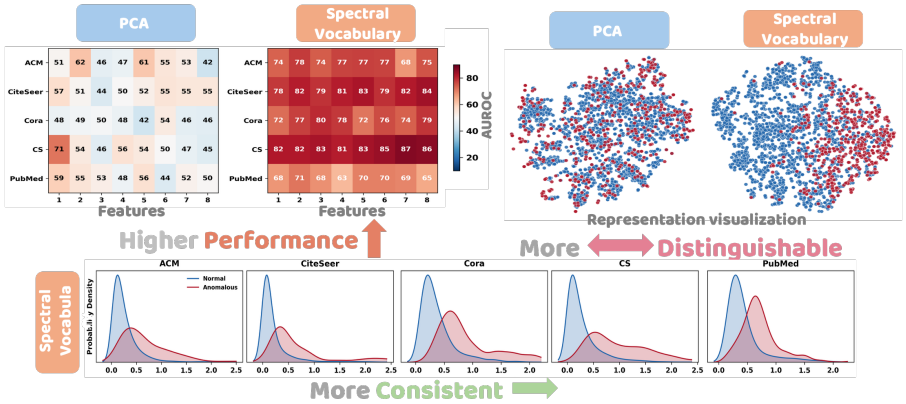


Figure 1: Comparison of spectral vocabulary with common GGAD representations. Top left: cross-dataset AUROC. Top right: representation visualization. Bottom: spectral vocabulary distributions of normal and anomalous nodes across datasets.

(2026a), providing a graph-relative description that is less dependent on the original feature coordinates or semantics. Our key insight is that, although graphs from different domains may have incompatible attribute spaces, their node features can still be characterized in a common way by how graph signals vary over the connectivity and how this variation is distributed across spectral frequencies. This enables us to define a shared spectral vocabulary whose coordinates have consistent graph-relative meaning across heterogeneous datasets.

Based on this insight, we propose **S**pectral **V**ocabulary for **G**raph **A**nomaly **D**etection (SV-GAD) a generalist graph anomaly detection framework for strict zero-shot transfer. The key idea of SV-GAD is to enable cross-domain transfer through a shared spectral vocabulary defined relative to graph structure. This representation provides consistent semantics across graphs despite differences in feature dimensionality and attribute meaning. On top of this vocabulary, SV-GAD leverages a spectral mixture-of-experts backbone that combines attention with learnable graph filters over multiple frequency bands, allowing the model to capture and adaptively integrate different spectral anomaly patterns across domains. SV-GAD is trained solely on source graphs and is directly applied to unseen target graphs without target labels, target-specific fine-tuning, or parameter adaptation. Our contributions are summarized as follows:

- **Spectral vocabulary for cross-domain graph representation.** We introduce a graph-structure-defined spectral vocabulary that transforms heterogeneous node attributes into transferable spectral descriptors. We further establish theoretical properties of the representation, including boundedness and scale invariance, providing a principled basis for cross-graph comparison.
- **Adaptive spectral modeling for domain heterogeneity.** We develop a spectral mixture-of-experts architecture to capture diverse anomaly patterns across domains. Multiple learnable spectral experts characterize different frequency behaviors, while adaptive routing allows the model to combine them according to the spectral characteristics of each graph.
- **Comprehensive evaluation under strict zero-shot transfer.** Extensive experiments across diverse graph domains demonstrate the transferability of the proposed spectral vocabulary and the effectiveness of SV-GAD. Our method achieves strong overall performance while scaling to graphs with substantially different structures, feature spaces, and anomaly characteristics.

2 Related Work

Graph Anomaly Detection. Node-level GAD identifies anomalous nodes from attributes and graph structure. Early methods rely on reconstruction (Ding et al., 2021; Fan et al., 2020), while later approaches use self-supervision and local context modeling. Spectral methods characterize anomalies through graph frequencies, including BWGNN (Tang et al., 2022), GHRN (Gao et al., 2023), and localized spectral formulations (Liu et al., 2026a). However, most methods require graph-specific training or adaptation (Tang et al., 2023), limiting transfer across heterogeneous graphs.

Generalist Graph Anomaly Detection. Recent work extends GAD toward generalist settings, where a single model transfers across unseen graphs. The key challenge is cross-domain mismatch in feature dimensions, semantics, and graph structure. Existing methods address this through feature alignment and prompting (Liu et al., 2024b; Niu et al., 2024), invariant or relational representations (Zhang et al., 2026; Liu et al., 2026b), and multi-expert modeling of heterogeneous anomaly geometries (Zhao et al., 2026). In contrast, we approach cross-domain transfer from a spectral perspective and introduce a bounded spectral graph vocabulary with graph-relative semantics, together with spectral experts that adapt to domain-dependent frequency patterns. A detailed comparison with existing GGAD methods is provided in Appendix B.

3 Graph Anomaly Detection in Spectral Domain

This section formulates graph anomaly detection from a spectral perspective. We first introduce the node-level GAD setting and the necessary background on spectral energy distributions, and then present our spectral energy formulation with boundedness guarantees.

Node-level GAD Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ denote an attributed graph, where $\mathcal{V} = \{v_i\}_{i=1}^N$ is the set of N nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set. $\mathbf{X} = \{x_1, \ldots, x_N\} \in \mathbb{R}^{N \times d}$ is the node feature matrix, where $\mathbf{x}_i \in \mathbb{R}^d$ denotes the feature vector of node v_i. Let $A \in \mathbb{R}^{N \times N}$ be the adjacency matrix of an undirected graph $\mathcal{G}$, and let D be its degree matrix, with $D_{ii} = \sum_j A_{ij}$. The normalized Laplacian is computed as follow, $\tilde{L} = I - D^{-1/2} A D^{-1/2}$. In this work, we focus on *node-level anomalies*. We denote the set of anomalous nodes as $\mathcal{V}_a \subseteq \mathcal{V}$ and the set of normal nodes as $\mathcal{V}_n \subseteq \mathcal{V}$, where $\mathcal{V}_a \cap \mathcal{V}_n = \emptyset$. Given the graph $\mathcal{G}$, node features $\mathbf{X}$, and labels for a subset of nodes, the objective is to learn a classifier that distinguishes anomalous nodes from normal ones.

Spectral Energy Distribution The normalized Laplacian matrix can be decomposed as $\tilde{L} = U \Lambda U^\top$, where $U = [u_1, \ldots, u_N]$ contains orthonormal eigenvectors and the corresponding eigenvalues satisfy $\lambda_1 \leq \cdots \leq \lambda_N$. Let $\hat{\mathbf{x}} = U^\top \mathbf{x}$ denote the graph Fourier transform of $\mathbf{x}$. Following prior work (Tang et al., 2022), the spectral energy, E, is expressed as $E = \frac{\mathbf{x}^\top \tilde{L} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$, which characterizes the anomalies in a right-shift behavior. However, this formulation requires substantial computation and does not show node-level information. Subsequent work uses MLP layers to obtain spectral energy for graph-level anomaly detection (Dong et al., 2024), and recent work (Liu et al., 2026a) extends to node-level anomaly detection by approximating spectral energy within local subgraphs. Following the previous way, we then define the local spectral energy E_i for one feature in the subgraph node $\mathbf{x}_{\mathcal{N}_i}$ as

$$E_i = \frac{\mathbf{x}_{\mathcal{N}_i}^\top \tilde{L}_{\mathcal{N}_i} \mathbf{x}_{\mathcal{N}_i}}{\mathbf{x}_{\mathcal{N}_i}^\top \mathbf{x}_{\mathcal{N}_i}} = \frac{\sum_{j \in \mathcal{N}(i)} (\tilde{x}_i - \tilde{x}_j)^2}{(\tilde{x}_i)^2 + \sum_{j \in \mathcal{N}(i)} (\tilde{x}_j)^2} = \frac{\eta_i}{\delta_i} \tag{1}$$

where $\tilde{L}_{\mathcal{N}_i}$ denotes the normalized Laplacian of the subgraph induced on $\mathcal{N}_i$. The normalized local spectral energy is obtained by applying the degree normalization $\tilde{x}_i = x_i / \sqrt{d_i}$ to the feature signal; we defer the full derivation to Appendix C.

Having a local energy E_i at every node gives a per-neighborhood view of smoothness, but it does not reflect how much each node should contribute globally. We extend this formulation to a weighted version and establish approximation error bounds:

Definition 3.1 (Weighted Spectral Energy)**.** *The weighted local spectral energy at node i and the aggregate weighted spectral energy over the graph are, respectively:*

$$\bar{E}_i = \frac{\delta_i \, E_i}{\sum_{i=1}^N \delta_i}, \qquad \bar{E} = \frac{\sum_{i=1}^N \delta_i \, E_i}{\sum_{i=1}^N \delta_i}, \tag{2}$$

where $\bar{E}_i$ quantifies node i's contribution to the global spectral energy as the local smoothness ratio $\tilde{E}_i$ weighted by the fraction $\delta_i / \sum_k \delta_k$ of total signal energy concentrated in its neighborhood.

For an undirected graph with no isolated nodes, the ratio between the degree-normalized and raw feature norms, $\rho = \frac{\|\tilde{\mathbf{x}}\|_2^2}{\|\mathbf{x}\|_2^2} = \frac{\sum_i x_i^2 / d_i}{\sum_i x_i^2} = \frac{\mathbf{x}^\top D^{-1} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$ is the Rayleigh quotient of the diagonal matrix D^{-1}, it is therefore bounded by the extremal degrees of the graph: $1/d_{\max} \leq \rho \leq 1/d_{\min}$.

Proposition 3.2. *For an undirected graph with no isolated nodes, the weighted aggregation of the local spectral energies is proportional to the global spectral energy, with proportionality constant* $2/(1+\rho)$ *determined by the graph's degree distribution.*

$$\bar{E} = \frac{\sum_i \eta_i}{\sum_i \delta_i} = \frac{2\,\mathbf{x}^\top \tilde{L}\,\mathbf{x}}{(1+\rho)\,\mathbf{x}^\top \mathbf{x}} = \frac{2}{1+\rho} \cdot E. \tag{3}$$

Proposition 3.2 establishes an exact equality between the global spectral energy E and the aggregate weighted local spectral energy $\bar{E}$, up to a constant determined by the graph's degree distribution. We refer the reader to Appendix D for the full derivation. We next present a formal bound to quantify the approximation error more precisely.

Corollary 3.3. *For any nonzero* $\mathbf{x} \in \mathbb{R}^N$*, the ratio between the weighted aggregation and global spectral energies is pinned to a degree-controlled interval:*

$$\frac{2d_{\min}}{d_{\min}+1} \leq \frac{\bar{E}}{E} \leq \frac{2d_{\max}}{d_{\max}+1} \tag{4}$$

Remark 3.4. At the sparse extreme ($d_i = 1$ for all i), $\bar{E} = E$ exactly; in the dense limit ($d_i \to \infty$), $\bar{E} \to 2E$. The weighted aggregation therefore interpolates between the global spectral energy and twice that value as graph density grows.

4 Transferable Pattern in Graph Anomaly Detection

Transferability is a fundamental challenge in building generalist or foundation models for GAD (Mao et al., 2024; Liu et al., 2024a). Graphs across domains differ greatly in both the dimensionality and the semantics of their node attributes, which makes it difficult to establish a shared feature space across domains. Existing GGAD methods address this mismatch in three main ways. The first applies dimensionality reduction to map attributes into a fixed dimension (Liu et al., 2024b). The second replaces raw attributes with universal representations, such as text embeddings (Tang et al., 2024). Both approaches provide a common feature space, but each has clear limitations. Dimensionality reduction aligns the dimensions but not the semantics, so the same dimension can still carry different meanings across datasets. Textual representations may not adequately characterize general graph data, especially when anomaly patterns depend on the interaction between nodes and local graph structure. The third approach uses predefined anomaly-related features. Such features have proven useful for GAD (Liu et al., 2026b), but they are usually low-dimensional and rely on hand-crafted statistics or domain knowledge.

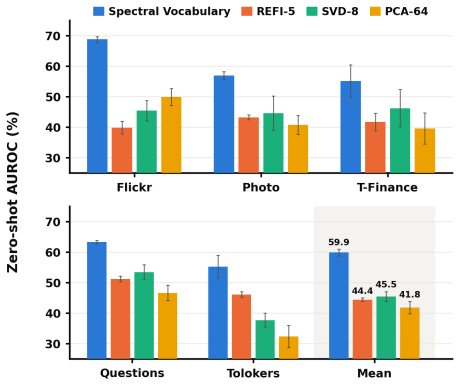


Figure 2: Zero-shot transfer of node representations. An MLP is trained on PubMed and Amazon and applied unchanged to five unseen target graphs.

These limitations call for a representation that provides a shared and semantically consistent vocabulary across graphs while preserving information relevant to GAD. Spectral energy offers a natural basis for such a representation. Defined over graph frequencies rather than raw feature coordinates, it provides a common description across heterogeneous feature spaces while characterizing how node signals vary with graph structure. Prior work has shown that graph anomalies can induce spectral energy shifts (Liu et al., 2026a) toward higher frequencies (Tang et al., 2022). Building on this observation, we introduce weighted spectral energy as a transferable vocabulary for GAD and investigate its cross-graph transferability in this section.

4.1 Spectral Vocabulary for Graph Anomaly Detection

Recent work suggests that an effective graph vocabulary should comprise a set of basic units that are transferable across datasets while preserving consistent semantics (Wang et al., 2024). Following this principle, we construct such a vocabulary from the weighted spectral energy of Eq. equation 2.

Definition 4.1 (Spectral graph vocabulary)**.** Given a graph $\mathcal{G}$, let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(C)} \in \mathbb{R}^N$ denote the C leading principal components of the column-standardized $\mathbf{X}$. The spectral profile of node v_i is defined as $\mathbf{s}_i = \left(\bar{E}_i^{(1)}, \ldots, \bar{E}_i^{(C)}\right) \in \mathbb{R}^C$, where $\bar{E}_i^{(c)}$ is the weighted spectral energy of v_i on channel $\mathbf{x}^{(c)}$. The spectral graph vocabulary of $\mathcal{G}$ is the matrix $\mathbf{S} := [\mathbf{s}_1, \ldots, \mathbf{s}_N]^\top \in \mathbb{R}^{N \times C}$.

Intuitively, the spectral vocabulary measures how strongly each channel's energy is concentrated at a node. Large values indicate strong local disagreement between a node and its neighbors relative to global signal power, giving a consistent graph-relative meaning across datasets. Each channel exactly decomposes global spectral energy up to the factor $2/(1+\rho)$ (Proposition 3.2), making the vocabulary informative for GAD: anomalies can shift energy toward higher frequencies (Tang et al., 2022) or, under distribution collapse, toward lower frequencies (Liu et al., 2026a). For the vocabulary to generalize across graphs, two additional properties are required: its entries must be invariant to how a dataset records its attributes (e.g., units, z-score, or min–max normalization), and they must remain bounded regardless of graph size and feature range.

Proposition 4.2 (Scale-free and bounded vocabulary)**.** *If $\mathcal{G}$ has no isolated nodes and every channel is nonzero, then:*

(i) Scale-free: *replacing a channel $\mathbf{x}^{(c)}$ by $a\mathbf{x}^{(c)}$, for any $a \neq 0$, leaves $\mathbf{S}$ unchanged.*

(ii) Bounded: *$S_{ic} \geq 0$, and every column of $\mathbf{S}$ sums to $\sum_i S_{ic} = \bar{E}^{(c)} < 4$, where $\bar{E}^{(c)}$ is the aggregate weighted spectral energy of channel $\mathbf{x}^{(c)}$. In particular, every entry lies in $[0, 4)$.*

We refer the reader to Appendix E for the proof. The proposition guarantees that the vocabulary is invariant to per-channel scaling and remain within fixed bounds. Therefore, the spectral vocabulary provides consistent semantics regardless of dataset-specific preprocessing or graph scale.

Empirical validation. To validate the transferability of the spectral vocabulary, we conduct a zero-shot experiment comparing it against common feature representations: PCA, SVD, and selected raw features. We train a simple MLP on top of each representation using only two source datasets and evaluate on unseen target graphs. As shown in Figure 2, the spectral vocabulary achieves an AUC-ROC approaching 60, substantially outperforming all alternative representations. This result highlights that the spectral vocabulary captures transferable structural patterns sufficient for cross-graph generalization, even with a minimal source set and a simple classifier.

5 METHOD

In this section, we introduce SV-GAD, our framework for generalist graph anomaly detection. SV-GAD first constructs a spectral graph vocabulary that maps heterogeneous graph attributes into a shared representation space while capturing node-level spectral variation across graphs. The resulting representation is then processed by a spectral mixture-of-experts backbone, which combines attention with learnable graph filters spanning multiple frequency bands and adaptively routes each node to different spectral experts. The model is trained on labeled source graphs and applied directly to unseen target graphs under a strict zero-shot setting.

5.1 PROBLEM FORMULATION

Generalist Graph Anomaly Detection. Let $\mathcal{T}_{\text{train}} = \{\mathcal{D}_{\text{train}}^{(1)}, \ldots, \mathcal{D}_{\text{train}}^{(n_s)}\}$ denote a collection of labeled source datasets, where each $\mathcal{D}_{\text{train}}^{(i)} = (\mathcal{G}_{\text{train}}^{(i)}, \mathbf{y}_{\text{train}}^{(i)})$ is drawn from an arbitrary domain. We aim to train a generalist graph anomaly detection model f on $\mathcal{T}_{\text{train}}$ that can identify anomalous nodes in any unseen target dataset $\mathcal{D}_{\text{test}}^{(j)} \in \mathcal{T}_{\text{test}}$, where $\mathcal{T}_{\text{test}} = \{\mathcal{D}_{\text{test}}^{(1)}, \ldots, \mathcal{D}_{\text{test}}^{(n_t)}\}$ is a disjoint collection of target datasets. The source and target datasets may come from different domains and follow different data distributions. Under our strict zero-shot setting, the model test in each target graph, without access to target labels, target-specific supervision, or fine-tuning.

5.2 SPECTRAL VOCABULARY FEATURE CONSTRUCTION

Common-width channels. The dimensions of graph datasets differ widely across domains (Table 2). We therefore first align every graph to a shared width C. For a dataset $\mathcal{D}^{(i)} \in \mathcal{T}_{\text{train}} \cup \mathcal{T}_{\text{test}}$

with features $\mathbf{X}^{(i)} \in \mathbb{R}^{N^{(i)} \times d^{(i)}}$, we compute

$$\mathbf{Z}^{(i)} = \mathrm{proj}\big(\mathbf{X}^{(i)}\big) \in \mathbb{R}^{N^{(i)} \times C}, \tag{5}$$

where proj is a linear map fitted to the graph's own features. For simplicity, we omit the superscript i in the following formulation. Any linear dimensionality reduction, such as PCA or SVD, can serve as proj. To enable meaningful projection, we first apply feature expansion for graphs with fewer attributes than the target channel dimension ($d^{(i)} < C$), and then project the expanded representation back to C channels.

Vocabulary entries. For each channel $\mathbf{z}^{(c)}$ we evaluate the local quantities of Eq. equation 1 on the degree-normalized signal $\tilde{\mathbf{z}}^{(c)} = D^{-1/2}\mathbf{z}^{(c)}$,

$$\eta_i^{(c)} = \sum_{j \in \mathcal{N}(i)} \big(\tilde{z}_i^{(c)} - \tilde{z}_j^{(c)}\big)^2, \qquad \delta_i^{(c)} = \big(\tilde{z}_i^{(c)}\big)^2 + \sum_{j \in \mathcal{N}(i)} \big(\tilde{z}_j^{(c)}\big)^2, \tag{6}$$

with $E_i^{(c)} = \eta_i^{(c)}/\delta_i^{(c)} \in [0, d_i + 1]$. The spectral graph vocabulary of node v_i on channel c is its weighted local spectral energy (Eq. equation 2). We rescaled it by the number of nodes N so that entries are comparable across graphs of different sizes:

$$S_{ic} = N\,\bar{E}_i^{(c)} = \frac{\eta_i^{(c)}}{\frac{1}{N}\sum_{k=1}^{N} \delta_k^{(c)}} = N\,E_i^{(c)}\,\frac{\delta_i^{(c)}}{\sum_{k=1}^{N} \delta_k^{(c)}}. \tag{7}$$

After computing the spectral profile for every node, we obtain the spectral graph vocabulary $\mathbf{S}$. To further enforce semantic alignment across datasets, we reorder the columns of both the projected features $\mathbf{Z}$ and the vocabulary $\mathbf{S}$ by their global spectral energy $\bar{E}$. To ensure a consistent channel ordering from smoothest to roughest. We then concatenate the two representations as $\mathbf{H} = [\mathbf{S} \,\|\, \mathbf{Z}]$. The projected features $\mathbf{Z}$ are included to preserve the basis information complementary to the spectral vocabulary; we validate the effectiveness of this concatenation in the ablation study (Section 6).

Anomalies do not live in one fixed frequency band. On many benchmarks they push a graph's spectral energy toward high frequencies, because an anomalous node differs sharply from its neighbors (Tang et al., 2022); on others they shift it toward low frequencies (Section 4.1). A filter fixed in advance therefore suits some graphs, but not others, and a detector for unseen graphs cannot know the right band beforehand. We therefore build the backbone around a mixture of spectral experts, which has a bank of filters that covers the spectrum from low to high frequencies, and a gate that chooses the appropriate filters for each node.

Multi-head attention. Each block first applies multi-head attention, so that every node gathers information from its neighbors:

$$\mathbf{H}^{\mathrm{a}} = \mathbf{H}^{(\ell-1)} + \mathrm{MHA}\big(\mathrm{LN}(\mathbf{H}^{(\ell-1)})\big), \tag{8}$$

where $\mathbf{H}^{(\ell-1)}$ is the input of block ℓ, LN denotes layer normalization, and MHA is a GAT layer () with redefined number of heads. Unlike the spectral filters below, which weight neighbors by the graph structure alone, attention computes its weights from the node states, so it can down-weight uninformative neighbors before filtering.

Mixture of spectral experts. The preceding representation is further processed by a MoE. We use a bank of $M + 1$ graph filters, consisting of a shared low-pass expert $\tilde{\beta}_0$ and M gated spectral experts. Specifically,

$$\begin{aligned} \mathrm{MoE}\big(\tilde{\mathbf{H}}^{\mathrm{a}}\big) &= \tilde{\beta}_0(\tilde{\mathbf{L}})\tilde{\mathbf{H}}^{\mathrm{a}}W_0 + \sum_{\nu=1}^{M} \boldsymbol{\alpha}_{:,\nu} \odot \tilde{\beta}_\nu(\tilde{\mathbf{L}})\tilde{\mathbf{H}}^{\mathrm{a}}W_\nu, \\ \boldsymbol{\alpha}_i &= \mathrm{softmax}\left(\mathrm{MLP}(\tilde{\mathbf{h}}_i^{\mathrm{a}})/\tau\right), \end{aligned} \tag{9}$$

where $\tilde{\mathbf{H}}^{\mathrm{a}} = \mathrm{LN}(\mathbf{H}^{\mathrm{a}})$, W_ν is the learnable projection of expert ν, and $\boldsymbol{\alpha}_i \in \mathbb{R}^M$ contains the routing weights for node v_i. Each expert filters the entire graph, while the node-wise gate determines how much of each expert output is retained for each node. Since the gate operates on the

attention-enhanced representation $\tilde{\mathbf{h}}_i^{\mathrm{a}}$, the routing decision incorporates both node and neighborhood information. The temperature τ controls the sharpness of the routing distribution, with smaller values producing more selective expert assignments.

Each spectral expert is parameterized as a learnable polynomial of the normalized Laplacian,

$$\tilde{\beta}_\nu(\tilde{\mathbf{L}}) = \sum_{k=0}^{K} c_{\nu,k}\, T_k(\tilde{\mathbf{L}} - \mathbf{I}), \tag{10}$$

where T_k denotes the k-th Chebyshev polynomial and $c_{\nu,k}$ are learnable coefficients. In the spectral domain, $\tilde{\beta}_\nu$ modulates each graph frequency λ by $\tilde{\beta}_\nu(\lambda)$, while the degree-$K$ polynomial restricts message propagation to at most K hops. We initialize the experts with Beta wavelets (Tang et al., 2022), ranging from low- to high-frequency responses, and allow training to adapt their spectral profiles. Finally, the MoE output in Eq. equation 9 is added to $\mathbf{H}^{\mathrm{a}}$ through a residual connection, followed by a feed-forward network with another residual connection to obtain $\mathbf{H}^{(\ell)}$.

Objective. A two-layer MLP maps the final node representation $\mathbf{h}_i^{(L)}$ to an anomaly score $\hat{y}_i$, where a larger value indicates a higher likelihood of being anomalous. For each source graph $\mathcal{G}_{\mathrm{train}}^{(i)}$, the model is optimized over its training nodes $\mathcal{V}_{\mathrm{train}}^{(i)}$ using a class-weighted binary cross-entropy loss. To account for class imbalance, the anomaly class is weighted by $\omega^{(i)}$, defined as the ratio of normal to anomalous training nodes.

We further regularize the routing gates during an initial warm-up stage using the node's spectral energy distribution. Let $\bar{\boldsymbol{\phi}}_i \in \mathbb{R}^M$ denote the fraction of node v_i's energy assigned to each routed spectral band. We encourage the gate $\boldsymbol{\alpha}_i^{(\ell)}$ at block ℓ to align with this distribution:

$$\mathcal{L}^{(g)} = \frac{1}{|\mathcal{V}_{\mathrm{tr}}^{(g)}|} \sum_{i \in \mathcal{V}_{\mathrm{tr}}^{(g)}} \Big[\mathrm{BCE}_{\omega^{(g)}}(\hat{y}_i, y_i) + \frac{\gamma}{L} \sum_{\ell=1}^{L} \mathrm{KL}\big(\bar{\boldsymbol{\phi}}_i \,\|\, \boldsymbol{\alpha}_i^{(\ell)}\big) \Big]. \tag{11}$$

The coefficient γ controls the strength of the spectral routing regularization and is set to zero after the warm-up stage.

6 Evaluation and Sensitivity Analysis

To evaluate the generalization ability of our model, we conduct extensive experiments on 15 datasets with 13 baselines under a strict zero-shot setting. Specifically, models have no access to target labels during training, validation, or inference. All hyperparameters, preprocessing procedures, and model-selection decisions are determined exclusively from source graphs and then applied unchanged to every unseen target graph, without target-specific tuning or adaptation. This protocol provides a rigorous evaluation of whether a model can transfer anomaly knowledge across heterogeneous graph domains rather than relying on information from the target data.

Dataset The experiments were conducted on widely used benchmark datasets spanning citation, social, e-commerce, and behavioral networks, ranging from small-scale (e.g., Citeseer) to large-scale (e.g., T-Social). Dataset statistics are summarized in Table 2.

Table 2: Dataset statistics. Shaded rows are used for training; the rest for testing.

Dataset	Domain	#Nodes	#Edges	#Feat.
ACM	Citation	16,484	74,073	8,337
Amazon	E-commerce	8,639	3,299,291	25
BlogCatalog	Social	5,196	172,779	8,189
Citeseer	Citation	3,327	5,077	3,703
Cora	Citation	2,708	5,802	1,433
Credit	Transaction	30,000	1,421,858	13
CS	Citation	18,333	83,993	6,805
Flickr	Social	7,575	241,301	12,047
PubMed	Citation	19,717	46,423	500
Questions	Social	48,921	153,540	301
T-Finance	Transaction	39,357	21,222,543	10
Tolokers	Behavioral	11,758	519,000	10
T-Social	Social	5,781,065	73,105,508	10
Weibo	Social	8,405	377,271	400
Yelp	E-commerce	45,954	3,846,979	32

Baselines We compare against thirteen baselines spanning two categories: (1) conventional graph models and GAD methods, which are GCN (Kipf & Welling, 2016), GAT (Veličković et al., 2017), GraphSAGE (Hamilton et al., 2017), BernNet (He et al., 2021), BWGNN (Tang et al., 2022), GHRN (Gao et al., 2023), and AMNet (); and (2) GGAD methods, including recent state-of-the-art approaches, which

Table 1: AUROC and AUPRC (%). Best results are bolded. OOM: out of memory.

Group	Method	ACM	CS	Cites.	Flickr	Yelp	Weibo	T-Fin.	Quest.	BlogC.	Tolok.	T-Soc.	Avg
						AUROC							
GAD	GCN (2017)	45.50	51.42	63.37	40.52	50.31	30.57	44.16	51.41	44.32	34.40	55.46	46.49
	GAT (2018)	70.90	69.48	59.77	66.19	47.53	58.61	51.39	55.41	67.82	42.07	40.77	57.27
	GraphSAGE (2017)	43.68	42.40	50.97	53.06	51.30	47.82	50.96	47.35	51.79	51.61	61.78	50.25
	BernNet (2021)	45.48	42.88	46.51	42.56	46.92	60.77	63.17	46.64	43.40	49.96	42.32	48.24
	AMNet (2022)	47.05	45.15	47.85	46.74	46.57	73.35	56.75	46.69	48.36	51.02	39.32	49.90
	BWGNN (2022)	53.52	48.81	47.42	62.59	46.95	56.62	64.43	43.81	57.76	48.89	41.32	52.01
	GHRN (2023)	49.35	45.23	48.51	61.47	46.41	64.88	55.07	44.13	58.06	51.97	OOM	52.51
GGAD	UNPrompt (2024)	71.48	71.82	59.01	68.69	57.10	82.10	39.75	50.75	68.52	48.63	43.77	60.15
	ARC (2024)	57.23	62.66	64.52	55.73	52.79	78.00	27.00	50.89	59.67	49.96	OOM	55.85
	AnomalyGFM (2025)	58.04	54.60	47.37	51.87	56.26	76.25	60.10	45.99	53.91	46.23	OOM	55.06
	IA-GGAD (2025)	72.65	83.68	84.44	70.74	54.28	**88.40**	25.75	56.14	70.10	43.49	OOM	64.97
	REFI-GAD (2026)	64.19	67.38	68.86	36.09	48.91	35.43	64.58	55.52	41.79	54.36	OOM	53.71
	GAD-MoRE (2026)	80.69	82.69	87.30	**73.88**	52.32	88.37	30.71	54.98	**74.01**	53.38	OOM	67.83
	SV-GAD (ours)	**92.06**	**96.56**	**93.76**	67.68	**57.14**	75.20	**67.23**	**56.60**	69.01	**58.46**	**75.82**	**73.59**
						AUPRC							
GAD	GCN (2017)	3.71	3.70	7.67	5.94	14.75	18.38	5.62	3.60	5.79	16.32	6.25	8.34
	GAT (2018)	22.66	16.47	7.88	19.89	14.08	30.98	5.57	3.93	20.52	19.84	2.36	14.93
	GraphSAGE (2017)	6.83	4.83	5.73	8.57	15.17	22.03	5.17	3.08	7.14	24.64	6.55	9.98
	BernNet (2021)	11.11	6.98	4.59	15.96	13.92	38.87	6.41	2.91	14.91	21.12	2.50	12.66
	AMNet (2022)	13.99	9.07	5.10	21.09	13.80	56.61	5.26	3.15	19.76	20.83	2.32	15.54
	BWGNN (2022)	16.47	13.03	4.95	24.00	13.72	28.21	7.13	2.81	21.27	21.37	2.51	14.13
	GHRN (2023)	14.54	9.07	5.14	23.63	13.52	33.90	5.86	2.89	22.01	22.61	OOM	15.32
GGAD	UNPrompt (2024)	15.02	13.37	7.38	24.28	19.14	48.00	3.57	3.22	31.46	22.73	2.47	17.33
	ARC (2024)	22.11	21.83	21.27	23.72	15.34	48.63	2.87	3.32	24.73	22.27	OOM	20.61
	AnomalyGFM (2025)	4.94	3.83	5.33	6.63	17.62	28.17	10.10	2.91	6.72	20.79	OOM	10.70
	IA-GGAD (2025)	36.66	31.66	33.52	**37.78**	15.32	**64.47**	2.81	3.96	31.95	20.97	OOM	27.91
	REFI-GAD (2026)	13.50	17.89	13.92	4.52	15.56	8.41	16.50	3.90	4.83	23.29	OOM	12.23
	GAD-MoRE (2026)	36.30	36.82	40.17	33.01	14.80	58.27	2.99	3.49	**36.95**	25.09	OOM	28.79
	SV-GAD (ours)	**46.20**	**57.44**	**65.97**	23.65	**19.24**	33.69	**16.56**	**4.03**	27.67	**29.45**	**10.45**	**30.40**

are UNPrompt (Niu et al., 2024), ARC (Liu et al., 2024b), AnomalyGFM (Qiao et al., 2025a), IA-GGAD (Zhang et al., 2026), REFI-GAD (Liu et al., 2026b), and GAD-MoRE (Zhao et al., 2026).

Experimental Setup. We report the standard GAD evaluation metrics, AUROC and AUPRC, averaged over ten random seeds for most datasets. For T-Social, we report results over five seeds due to its large scale and computational cost. Each baseline is trained once on $\mathcal{T}_{\text{train}}$ and then evaluated on all target graphs in a strict zero-shot setting. Hyperparameters are selected via random search using only the source training split and are fixed across all target graphs, with no target-specific tuning.

Overall Performance. Table 1 summarizes the zero-shot anomaly detection performance across 11 target datasets. SV-GAD achieves the best AUROC on 8/11 datasets and the best AUPRC on 8/11 datasets. Based on the reported average performance, SV-GAD improves over the second-best method by 8.49% in AUROC and 5.59% in AUPRC. The gains are particularly pronounced on challenging cross-domain targets: compared with the strongest baseline on each dataset, SV-GAD improves AUROC by +12.88 points on CS and +11.37 points on ACM, while improving AUPRC by +25.80 points on CiteSeer. Notably, SV-GAD also scales successfully to the large-scale T-Social dataset, where several competing GGAD methods run out of memory. This demonstrates that SV-GAD maintains both strong transferability and practical scalability across different domain graphs. Although SV-GAD does not achieve the best result on Flickr, Weibo, and BlogCatalog, its consistently strong performance across the remaining targets supports the effectiveness of the spectral vocabulary and adaptive spectral experts for zero-shot transfer.

Table 3: AUROC (%) by target domain.

Domain	SV-GAD	IA-GGAD	ARC	AnomalyG.	GIN
Citation	**82.57** ↑+0.1	82.48	63.81	50.54	39.70
Social	59.25 ↓5.3	**64.58**	61.20	50.04	40.08
Transaction	**62.39** ↑10.1	32.35	37.40	39.83	52.31
Behavioral	**55.97** ↑3.7	44.47	53.11	42.90	52.27

Cross-Domain Transfer. We further evaluate cross-domain generalization by training all methods exclusively on e-commerce graphs and testing on unseen citation, social, transaction, and behavioral

graphs. As shown in Table 3, SV-GAD achieves the highest overall AUROC and leads in three of the four target domains—citation, transaction, and behavioral—including a margin of more than 10 points on transaction graphs. These results indicate that the proposed spectral vocabulary captures anomaly patterns that remain transferable even under substantial shifts in feature semantics and graph structure. Performance is weaker on social graphs, where the anomaly distribution may differ more substantially from the e-commerce source domain, leading to changes in spectral characteristics that are difficult to capture from a single source domain. Overall, SV-GAD demonstrates strong cross-domain transfer and robustness across heterogeneous graph types.

Hyperparameter Sensitivity. We study the sensitivity to the alignment width C and the number of routed experts M. Increasing C generally improves performance, with the average AUROC increasing from 84.1 at C=8 to 87.3 at C=64, mainly driven by the improvement on ACM (84.2 → 93.6). CS and Citeseer remain relatively stable across different values of C. For the number of experts, M shows a dataset-dependent trade-off: increasing M benefits CS and ACM, while Flickr performs better with fewer experts. Overall, the average performance remains relatively stable across different choices of M, indicating that the model is not highly sensitive to this hyperparameter.

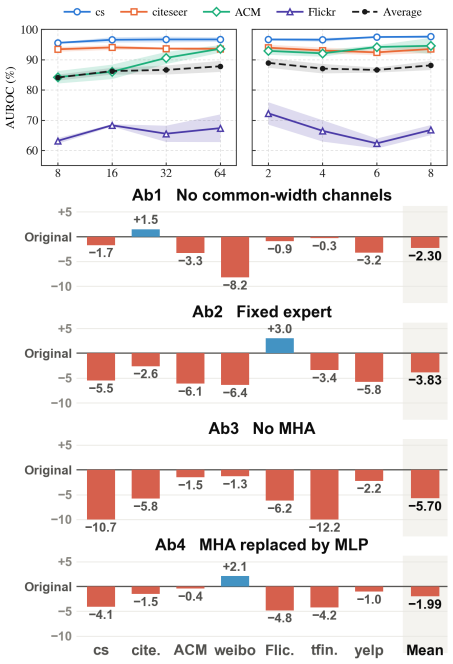


Figure 3: **Sensitivity and ablation.** Top: AUROC under different alignment widths C and numbers of experts M. Bottom: AUROC changes across ablated variants.

Component Analysis. We conduct four ablation studies to assess the contribution of the main components of SV-GAD: removing the common-width channels **Z**, replacing the mixture of spectral experts with a single high-pass expert, removing MHA, and replacing MHA with a two-layer MLP. As shown in Fig. 3, all four variants decrease the mean AUROC, confirming that each component contributes to the full model. Removing MHA causes the largest degradation, while replacing it with an MLP partially recovers performance but remains consistently worse, indicating that attention-based interactions are important for aggregating neighboring representations. Removing **Z** also reduces performance, showing the benefit of the shared aligned representation, while using only a single high-pass expert weakens performance across most datasets, supporting the need for multiple spectral experts to capture diverse anomaly frequency patterns.

7 Conclusion and Future Direction

In this work, we introduced a spectral vocabulary for generalist graph anomaly detection, providing a shared graph-relative representation across datasets with heterogeneous feature spaces and structures. Based on this vocabulary, we developed SV-GAD, which combines multiple spectral experts to capture diverse anomaly patterns and enables direct transfer to unseen graphs without target-label adaptation. Extensive experiments across heterogeneous graph domains demonstrate that SV-GAD achieves strong and scalable performance under strict zero-shot transfer, supporting spectral representations as a promising basis for transferable graph anomaly detection. While this work focuses on node-level anomalies, an important direction for future work is to extend the spectral vocabulary toward a unified GGAD framework that supports edge-level, and graph-level anomaly detection.

STATEMENT OF AI USE

AI tools supported manuscript organization, drafting, figure design, and implementation checks. Existing research artifacts supply the empirical claims; missing measurements are marked explicitly. The authors remain responsible for the final manuscript and its evidence.

REFERENCES

Kaize Ding, Qinghai Zhou, Hanghang Tong, and Huan Liu. Few-shot network anomaly detection via cross-network meta-learning. In *Proceedings of the web conference 2021*, pp. 2448–2456, 2021.

Xiangyu Dong, Xingyi Zhang, and Sibo Wang. Rayleigh quotient graph neural networks for graph-level anomaly detection. In *International Conference on Learning Representations*, volume 2024, pp. 53937–53955, 2024.

Mingjiang Duan, Tongya Zheng, Yang Gao, Gang Wang, Zunlei Feng, and Xinyu Wang. Dga-gnn: Dynamic grouping aggregation gnn for fraud detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 11820–11828, 2024.

Haoyi Fan, Fengbin Zhang, and Zuoyong Li. Anomalydae: Dual autoencoder for anomaly detection on attributed networks. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5685–5689. IEEE, 2020.

Yuan Gao, Xiang Wang, Xiangnan He, Zhenguang Liu, Huamin Feng, and Yongdong Zhang. Addressing heterophily in graph anomaly detection: A perspective of graph spectrum. In *Proceedings of the ACM web conference 2023*, pp. 1528–1538, 2023.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Mingguo He, Zhewei Wei, Hongteng Xu, et al. Bernnet: Learning arbitrary graph spectral filters via bernstein approximation. *Advances in neural information processing systems*, 34:14239–14251, 2021.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Ranran Li, Zhaowei Liu, Yuanqing Ma, Dong Yang, and Shuaijie Sun. Internet financial fraud detection based on graph learning. *IEEE Transactions on Computational Social Systems*, 10(3): 1394–1401, 2022.

Hao Liu, Jiarui Feng, Lecheng Kong, Ningyue Liang, Dacheng Tao, Yixin Chen, and Muhan Zhang. One for all: Towards training one graph model for all classification tasks. In *International conference on learning representations*, volume 2024, pp. 20188–20210, 2024a.

Yilin Liu, Hongchao Zhang, Taylor T Johnson, Ahmad F Taha, and Meiyi Ma. Modeling spectral energy shifts in spatio-temporal graph anomaly detection. *arXiv preprint arXiv:2606.00304*, 2026a.

Yixin Liu, Shiyuan Li, Yu Zheng, Qingfeng Chen, Chengqi Zhang, and Shirui Pan. Arc: A generalist graph anomaly detector with in-context learning. *Advances in Neural Information Processing Systems*, 37:50772–50804, 2024b.

Yujing Liu, Yixin Liu, Yu Zheng, Alan Wee-Chung Liew, Xiaofeng Cao, and Shirui Pan. Rethinking feature alignment in generalist graph anomaly detection: A relational fingerprint-based approach. *arXiv preprint arXiv:2605.25429*, 2026b.

Haitao Mao, Zhikai Chen, Wenzhuo Tang, Jianan Zhao, Yao Ma, Tong Zhao, Neil Shah, Mikhail Galkin, and Jiliang Tang. Position: Graph foundation models are already here. *arXiv preprint arXiv:2402.02216*, 2024.

Chaoxi Niu, Hezhe Qiao, Changlu Chen, Ling Chen, and Guansong Pang. Zero-shot generalist graph anomaly detection with unified neighborhood prompts. *arXiv preprint arXiv:2410.14886*, 2024.

Hezhe Qiao, Chaoxi Niu, Ling Chen, and Guansong Pang. Anomalygfm: Graph foundation model for zero/few-shot anomaly detection. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 2326–2337, 2025a.

Hezhe Qiao, Hanghang Tong, Bo An, Irwin King, Charu Aggarwal, and Guansong Pang. Deep graph anomaly detection: A survey and new perspectives. *IEEE Transactions on Knowledge and Data Engineering*, 37(9):5106–5126, 2025b.

Jiabin Tang, Yuhao Yang, Wei Wei, Lei Shi, Lixin Su, Suqi Cheng, Dawei Yin, and Chao Huang. Graphgpt: Graph instruction tuning for large language models. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, pp. 491–500, 2024.

Jianheng Tang, Jiajin Li, Ziqi Gao, and Jia Li. Rethinking graph neural networks for anomaly detection. In *International conference on machine learning*, pp. 21076–21089. PMLR, 2022.

Jianheng Tang, Fengrui Hua, Ziqi Gao, Peilin Zhao, and Jia Li. Gadbench: Revisiting and benchmarking supervised graph anomaly detection. *Advances in Neural Information Processing Systems*, 36:29628–29653, 2023.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Zehong Wang, Zheyuan Zhang, Nitesh V Chawla, Chuxu Zhang, and Yanfang Ye. Gft: Graph foundation model with transferable tree vocabulary. *Advances in neural information processing systems*, 37:107403–107443, 2024.

Xiong Zhang, Zhenli He, Changlong Fu, and Cheng Xie. Ia-ggad: Zero-shot generalist graph anomaly detection via invariant and affinity learning. *Advances in Neural Information Processing Systems*, 38:83728–83760, 2026.

Xinyu Zhao, Qingyun Sun, Jiayi Luo, Xingcheng Fu, and Jianxin Li. Zero-shot generalizable graph anomaly detection with mixture of riemannian experts. *arXiv preprint arXiv:2602.06859*, 2026.

A NOTAIONS

Table 4: Notation.

Symbol	Meaning	Symbol	Meaning
$\mathcal{G}$	Graph	$\mathcal{V}$	Set of nodes
$\mathcal{E}$	Set of edges	N	The number of nodes
$\mathbf{X}$	Node feature matrix	A	Adjacency matrix
D	Degree matrix	$\tilde{L}$	Normalized Laplacian matrix
U	Eigenvector matrix	Λ	Eigenvalues
$\mathcal{N}_i$	Local subgraph of node v_i	E	Global spectral energy
E_i	Local spectral energy at node v_i	$\bar{E}_i$	Weighted local spectral energy at node v_i
$\bar{E}$	Aggregate weighted spectral energy	ρ	Degree-normalized to raw feature norm ratio
C	The number of channels	$\mathbf{Z}$	Projected features
$\mathbf{S}$	Spectral graph vocabulary	$\mathbf{H}$	Node representation
M	The number of spectral experts	$\tilde{\beta}_\nu$	Spectral expert ν
$\boldsymbol{\alpha}_i$	Routing weights of node v_i	$\mathcal{T}_{\text{train}}$	Source datasets
$\mathcal{T}_{\text{test}}$	Target datasets	$\mathcal{L}$	Training loss

B MORE DETAIL IN RELATED WORK

Graph Anomaly Detection. Node-level graph anomaly detection (GAD) aims to identify nodes whose attributes, structural roles, or local interactions deviate from normal patterns. Early unsupervised methods commonly detect anomalies through reconstruction errors, using graph autoencoders or joint reconstruction of node attributes and topology (Ding et al., 2021; Fan et al., 2020). Later approaches increasingly adopt self-supervised learning, contrastive objectives, and local-context modeling to obtain more discriminative anomaly representations. Another line of work studies GAD from the graph spectral perspective. BWGNN (Tang et al., 2022) associates anomalies with high-frequency graph signals and introduces Beta-wavelet filters to capture frequency-specific patterns, while GHRN (Gao et al., 2023) leverages spectral information to mitigate anomaly-induced heterophily. More recent work develops localized spectral formulations that characterize node-level spectral energy and improve scalability to large graphs (Liu et al., 2026a). Despite strong performance, most conventional GAD methods are optimized for a single graph and typically require graph-specific training, hyperparameter tuning, or adaptation (Tang et al., 2023), which limits their applicability to unseen and heterogeneous graph domains.

Generalist Graph Anomaly Detection. Recent work extends GAD toward generalist settings, where a single detector is expected to transfer across previously unseen graphs. This setting is substantially more challenging because graphs may differ in feature dimensionality, attribute semantics, topology, and anomaly distributions. Existing methods address these discrepancies from several perspectives. ARC (Liu et al., 2024b) combines cross-graph feature alignment with in-context learning, while UNPrompt (Niu et al., 2024) enables zero-shot transfer through normalized feature alignment and unified neighborhood prompting. IA-GGAD (Zhang et al., 2026) explicitly models feature-space and structural shifts through invariant and affinity learning. ReFi-GAD (Liu et al., 2026b) constructs relational fingerprints as transferable anomaly descriptors, whereas GAD-MoRE (Zhao et al., 2026) employs multiple Riemannian experts to model heterogeneous anomaly geometries across graphs. These approaches demonstrate the potential of generalist GAD, but largely rely on learned alignment, predefined relational statistics, or geometry-specific representations. In contrast, we study cross-domain transfer from a spectral perspective and introduce a bounded spectral graph vocabulary whose dimensions have consistent graph-relative semantics across datasets. Building on this vocabulary, multiple spectral experts capture domain-dependent frequency patterns and enable direct transfer to unseen graphs without target-label adaptation.

C Derivation of equation equation 1.

We have $\tilde{L}_{\mathcal{N}_i} = I - D_{\mathcal{N}_i,f}^{-1/2} A_{\mathcal{N}_i,f} D_{\mathcal{N}_i,f}^{-1/2}$. Substituting $\tilde{L}_{\mathcal{N}_i}$ into equation 1 for neighborhood $\mathcal{N}_i$ and expanding via matrix multiplication, we obtain $\tilde{E}_i = E_{num}/E_{den}$, where the numerator and denominator are given by

$$E_{num} = \sum_{j \in \mathcal{N}(i)} w_{ij} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2 = \eta_i, \;\; E_{den} = \left(\frac{x_i}{\sqrt{d_i}} \right)^2 + \sum_{j \in \mathcal{N}(i)} \left(\frac{x_j}{\sqrt{d_j}} \right)^2 = \delta_i$$

D Proof of Proposition 3.2

The δ_i-weighting cancels the per-node ratios and reduces $\bar{E}$ to $\sum_i \eta_i / \sum_i \delta_i$:

$$\bar{E} \;=\; \frac{\sum_i \delta_i \cdot E_i}{\sum_i \delta_i} \;=\; \frac{\sum_i \delta_i \cdot (\eta_i/\delta_i)}{\sum_i \delta_i} \;=\; \frac{\sum_i \eta_i}{\sum_i \delta_i}.$$

We evaluate the numerator and the denominator separately.

Numerator. Expanding η_i and converting the ordered-pair sum into an undirected edge sum gives

$$\sum_i \eta_i \;=\; \sum_i \sum_{j \in \mathcal{N}(i)} (\tilde{x}_i - \tilde{x}_j)^2 \;=\; 2 \sum_{(i,j) \in \mathcal{E}} (\tilde{x}_i - \tilde{x}_j)^2 \;=\; 2\big(\tilde{\mathbf{x}}^\top D\, \tilde{\mathbf{x}} - \tilde{\mathbf{x}}^\top A\, \tilde{\mathbf{x}}\big) \;=\; 2\, \tilde{\mathbf{x}}^\top L\, \tilde{\mathbf{x}}.$$

Substituting $\tilde{\mathbf{x}} = D^{-1/2}\mathbf{x}$ and using $D^{-1/2} L\, D^{-1/2} = \tilde{L}$,

$$\tilde{\mathbf{x}}^\top L\, \tilde{\mathbf{x}} \;=\; \mathbf{x}^\top \big(D^{-1/2} L\, D^{-1/2}\big)\mathbf{x} \;=\; \mathbf{x}^\top \tilde{L}\, \mathbf{x}$$

so that $\sum_i \eta_i = 2\, \mathbf{x}^\top \tilde{L}\, \mathbf{x}$.

Denominator. Expanding δ_i and swapping the order of summation in the second term,

$$\sum_i \delta_i \;=\; \sum_i \tilde{x}_i^2 + \sum_i \sum_{j \in \mathcal{N}(i)} \tilde{x}_j^2 \;=\; \sum_i \tilde{x}_i^2 + \sum_j \tilde{x}_j^2 \underbrace{\sum_{i \in \mathcal{N}(j)} 1}_{d_j} \;=\; \sum_i \tilde{x}_i^2 + \sum_i d_i\, \tilde{x}_i^2.$$

Using $\tilde{\mathbf{x}} = D^{-1/2}\mathbf{x}$, the identity $\sum_i d_i\, \tilde{x}_i^2 = \sum_i x_i^2$ holds, so

$$\sum_i \delta_i \;=\; \|\tilde{\mathbf{x}}\|_2^2 + \|\mathbf{x}\|_2^2 \;=\; (1+\rho)\, \mathbf{x}^\top \mathbf{x},$$

where $\rho \;=\; \|\tilde{\mathbf{x}}\|_2^2 / \|\mathbf{x}\|_2^2$.

Combining. Dividing the two expressions yields

$$\bar{E} \;=\; \frac{2\, \mathbf{x}^\top \tilde{L}\, \mathbf{x}}{(1+\rho)\, \mathbf{x}^\top \mathbf{x}} \;=\; \frac{2}{1+\rho} \cdot \frac{\mathbf{x}^\top \tilde{L}\, \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \;=\; \frac{2}{1+\rho}\, E$$

E Proof of PROPOSITION 4.2

Proof. Fix a channel c and write $\mathbf{x} = \mathbf{x}^{(c)}$, dropping the superscript (c) from η_i, δ_i, and $\bar{E}_i$.

(i) Replacing $\mathbf{x}$ by $a\mathbf{x}$ replaces its degree-normalized version $\tilde{\mathbf{x}} = D^{-1/2}\mathbf{x}$ by $a\tilde{\mathbf{x}}$. Since η_i and every δ_k are quadratic in $\tilde{\mathbf{x}}$, all of them are multiplied by a^2, which cancels in $\bar{E}_i = \eta_i / \sum_k \delta_k$. Hence every entry $S_{ic} = \bar{E}_i$ is unchanged.

(ii) Each entry $S_{ic} = \eta_i / \sum_k \delta_k$ is nonnegative, since $\eta_i \geq 0$ and $\sum_k \delta_k > 0$. Summing Eq. equation 1 over all nodes gives $\sum_i \eta_i \;=\; 2\, \tilde{\mathbf{x}}^\top (D - A) \tilde{\mathbf{x}}$ and $\sum_i \delta_i \;=\; \tilde{\mathbf{x}}^\top (D + I) \tilde{\mathbf{x}}$. The numerator satisfies

$$\tilde{\mathbf{x}}^\top (D - A) \tilde{\mathbf{x}} = \sum_{\{i,j\} \in \mathcal{E}} (\tilde{x}_i - \tilde{x}_j)^2 \leq 2 \sum_{\{i,j\} \in \mathcal{E}} (\tilde{x}_i^2 + \tilde{x}_j^2) = 2\, \tilde{\mathbf{x}}^\top D\, \tilde{\mathbf{x}},$$

and, since $d_i \le d_{\max}$ for every node,

$$D \preceq \frac{d_{\max}}{d_{\max}+1}(D+I).$$

Therefore,

$$0 \le \sum_i S_{ic} = \frac{2\,\tilde{\mathbf{x}}^\top (D-A)\tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^\top (D+I)\tilde{\mathbf{x}}} \le \frac{4d_{\max}}{d_{\max}+1} < 4,$$

so every entry of $\mathbf{S}$ lies in $[0, 4)$. □

F Additional Experiment Details

F.1 Experimental Setup

All the experiments were run in a single workstation, which is equipped with an AMD Ryzen Threadripper PRO 7975WX processor (32 cores, 64 threads), 128 GB of DDR5 RAM, and an NVIDIA RTX 6000 Ada Generation GPU with 48 GB of VRAM.

F.2 Baseline

We compare SV-GAD with thirteen baselines spanning conventional graph/GAD methods and recent generalist graph anomaly detection methods.

- **GCN** (Kipf & Welling, 2016): GCN performs neighborhood aggregation through graph convolution to learn node representations.
- **GAT** (Veličković et al., 2017): GAT uses attention mechanisms to assign different importance weights to neighboring nodes during message passing.
- **GraphSAGE** (Hamilton et al., 2017): GraphSAGE learns inductive node representations by sampling and aggregating features from local neighborhoods.
- **BernNet** (He et al., 2021): BernNet learns flexible spectral graph filters using Bernstein polynomial approximations over the graph spectrum.
- **AMNet**: AMNet models anomaly-related graph signals through adaptive multi-frequency filtering.
- **BWGNN** (Tang et al., 2022): BWGNN models graph anomalies from a spectral perspective and uses Beta-wavelet filters to capture high-frequency anomaly signals.
- **GHRN** (Gao et al., 2023): GHRN leverages spectral information to mitigate anomaly-induced heterophily and improve anomaly discrimination.
- **UNPrompt** (Niu et al., 2024): UNPrompt enables zero-shot graph anomaly detection through normalized feature alignment and unified neighborhood prompting.
- **ARC** (Liu et al., 2024b): ARC performs one-for-all graph anomaly detection through cross-graph feature alignment and in-context learning.
- **AnomalyGFM** (Qiao et al., 2025a): AnomalyGFM learns transferable normal and anomalous representations to support zero-shot and few-shot anomaly detection across graphs.
- **IA-GGAD** (Zhang et al., 2026): IA-GGAD addresses cross-graph feature and structural shifts through invariant representation and affinity learning.
- **REFI-GAD** (Liu et al., 2026b): REFI-GAD constructs relational fingerprints to obtain transferable anomaly representations across heterogeneous graphs.
- **GAD-MoRE** (Zhao et al., 2026): GAD-MoRE employs multiple geometry-aware experts to model heterogeneous anomaly patterns across graph domains.

F.3 Metrics

AUPRC (Area Under the Precision-Recall Curve) AUPRC is used in a classification model to show the relationship between precision and recall at different threshold levels, which computes the area beneath the Precision-Recall curve.

AUROC (Area Under the Receiver Operating Characteristic Curve) AUROC evaluates a model's ability to distinguish the positive and negative classes by measuring the area under the ROC curve. The ROC curve plots the true positive rate against the false positive rate for varying decision thresholds. The ROC approach to 1 shows it has perfect ability to distinguish the different classes.