TAS-EGNN: Task-Aware Spectral Ego-Graphs for Efficient GNNs-Based Classification

Mebarka Allaoui[1] **Rachid Hedjam**[1] **Sonia Gupta**[2]
[1]Bishop's University, Sherbrooke, QC, Canada [2]AI Garage, Mastercard, Gurugram, India

Abstract

Graph Neural Networks (GNNs) achieve strong accuracy but remain costly to train on large graphs and in resource-constrained settings. Coreset selection mitigates this by training on a compact, representative node subset, yet many existing methods rely on expensive spectral routines or bilevel and iterative optimizations. We propose a Task-Aware Spectral Ego-Graph Neural Network (TAS-EGNN) that scores nodes within lightweight ego-graphs by combining (i) local spectral complexity, (ii) predictive uncertainty, and (iii) supervised error signals, followed by a greedy coverage step to avoid redundancy. TAS-EGNN circumvents heavy optimization, using only local spectra (or moment proxies) and a single model forward pass to obtain task signals. We evaluate TAS-EGNN across several benchmark tasks: citation networks, social networks, product networks, and graph-based bank transaction fraud detection. The last task, in particular, underscores the algorithm's effectiveness in anomaly detection for highly imbalanced settings. TAS-EGNN matches or surpasses state-of-the-art reduction baselines, across *budgets* (i.e., the allowed size of the selected training subset, controlled via the coreset ratio), including condensation, coarsening, and ego-graph selection, while delivering substantial time and memory savings. Time and memory profiling show that TAS-EGNN tracks the lower envelope among structure-aware methods and scales to large graphs, whereas several other works reach OOT/OOM. These results indicate that efficiently encoded task-aware structural priors enable robust, scalable coreset selection for both standard node classification and fraud detection. The source code is available on GitHub.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

1 INTRODUCTION

Graph Neural Networks (GNNs) have become foundational tools for modeling relational data, showing state-of-the-art performance in node classification, link prediction, and graph-level tasks across domains such as social networks, molecular biology, and recommendation systems Kipf and Welling (2016); Veličković et al. (2018); Xu et al. (2018); Hamilton et al. (2017); Zhang et al. (2019); Gilmer et al. (2017). Despite these successes, a persistent challenge remains: training efficiency and scalability on large-scale graphs. This bottleneck has driven increased interest in coreset selection techniques Sener and Savarese (2017); Mirzasoleiman et al. (2020); Killamsetty et al. (2021), which aim to identify representative subsets of training nodes (core node set), enabling faster training while maintaining model performance.

Coreset selection in graphs is particularly challenging due to non-i.i.d. node dependencies (a node's signal depends on its neighbors) and the need to retain structural information. Several strategies have been proposed, including diversity-oriented coverage of the graph Navlakha et al. (2008); Dubey et al. (2018), supervision-aware selection using gradients or misclassification Tong et al. (2025); Killamsetty et al. (2021); Nagaraj et al. (2025), and spectral objectives that preserve structural properties Jin et al. (2020); Ding et al. (2024). Spectral methods seek to preserve global graph structure but typically require full eigen-decomposition, limiting their applicability to small graphs Jin et al. (2020); Ding et al. (2024). However, existing approaches face several fundamental limitations. First, strategies relying on global graph operations or exhaustive gradient evaluations are computationally prohibitive for large-scale scenarios. Second, methods that focus solely on either structure or task-specific signals (such as entropy or misclassification) often fail to capture the

nuanced interplay between topology and learning difficulty. Third, most approaches degrade substantially when applied under strict coreset budget constraints, which are a common requirement in real-world applications such as graph streaming, active learning, or federated learning Duan et al. (2022); Waikhom and Patgiri (2023). Taken together, these gaps leave open the problem of selecting training subsets that *simultaneously* satisfy three practical desiderata: (1) identifying regions of high predictive uncertainty, which are crucial in safety-critical tasks such as fraud and intrusion detection Alsaadi et al. (2020); (2) capturing nodes near decision boundaries or frequently misclassified, which are often essential for improving model generalization; and (3) ensuring diversity among selected subgraphs to avoid redundancy, especially when subgraphs exhibit heavy overlap in sparse or hierarchical networks. These aspects are vital for building effective and compact coresets in real-world graph learning scenarios, yet have remained largely unaddressed in prior work.

In response to these limitations, we propose TAS-EGNN, a Task-Aware Spectral based on Ego-Graph selection for efficient GNN training. At its core is a scoring function that fuses three signals computed on small r-hop ego-graphs: (1) *local spectral complexity* via spectral entropy to capture structural heterogeneity; (2) *predictive entropy* to quantify model uncertainty; and (3) a *misclassification indicator* to emphasize error-prone nodes. To remain scalable, TAS-EGNN avoids global graph operations: we extract and score ego-graphs locally, obtain task signals from a single forward pass of the backbone model, and select centers with a greedy coverage rule that maximizes marginal gain while penalizing overlap, yielding informative yet non-redundant subsets. Our method addresses three critical gaps in the field: i) it eliminates the need for expensive global computations such as eigen-decompositions or full backpropagation across the entire graph, enabling application to large-scale datasets; ii) it maintains strong coreset coverage and discriminative power under low budget regimes by integrating structure and task difficulty signals; and iii) it directly couples structural richness with task relevance, capturing synergistic effects often overlooked by prior methods. By integrating these elements into a unified, efficient, and localized framework, TAS-EGNN advances the state of the art in graph coreset selection, offering a scalable and theoretically grounded approach to learning from fewer nodes. Beyond standard citation, social, and product benchmarks, our objective is to test whether TAS-EGNN remains effective under high-stakes, severely imbalanced settings such as transaction fraud detection. Accordingly, we evaluate the performance of TAS-EGNN against the baseline techniques on benchmark datasets in terms of node classification and computational efficiency. We summarize our contributions as follows:

- We formulate a task-aware scoring function that balances local structure, uncertainty, and misclassification to rank nodes for coreset inclusion.
- We design a scalable, greedy algorithm for subset selection based on ego-graph coverage, suitable for inductive and transductive settings.
- We empirically validate TAS-EGNN on a set of benchmark datasets belonging to several tasks, which are citation networks, social networks, product networks, and graph-based bank transaction fraud detection, demonstrating superior performance compared to existing coreset and spectral selection strategies.
- We provide a comprehensive efficiency study (log-scale time and peak memory) across all datasets, showing that TAS-EGNN consistently tracks the lower envelope among structure-aware baselines and scales to large graphs where several competitors hit OOT/OOM.

The remainder of the paper is organized as follows. Section 2 introduces the TAS-EGNN framework. Section 3 presents empirical evaluations and ablation studies. Section 4 concludes the paper. The detailed related work discussion is provided in App. A.

2 TASK-AWARE SPECTRAL EGO GRAPH NEURAL NETWORK

In this section, we introduce **TAS-EGNN**, a task-aware coreset selection framework for efficient GNN training. Notice that all symbols used in this section are defined in Tab. 1.

Table 1: Summary of notations.

Symbol	Description
$\mathcal{G} = (V, E, X, Y)$	Input graph with nodes, edges, features, and labels
$A \in \{0,1\}^{n \times n}$	Adjacency matrix of the graph
$D \in \mathbb{R}^{n \times n}$	Degree matrix; $D_{ii} = \sum_j A_{ij}$
L, L_v	Normalized Laplacian on $\mathcal{G}$ and on $\mathcal{E}_v$
$X \in \mathbb{R}^{n \times d}$	Node feature matrix
$Y \in \{1, \ldots, C\}^n$	Ground-truth labels
$f_\theta(A, X)$	GNN with parameters θ
f_v	Logits predicted for node v
$\hat{y}_v$	$\arg\max_c f_v^{(c)}$
$p_v^{(c)}$	$\mathrm{softmax}(f_v)^{(c)}$
$\mathcal{E}_v$	r-hop ego-graph around v
A_v, X_v	Adjacency and features on $\mathcal{E}_v$
λ_i	Eigenvalue i of L
$\mathcal{H}_{\mathrm{spec}}(\mathcal{E}_v)$	Spectral entropy of $\mathcal{E}_v$
$\mathcal{H}_{\mathrm{pred}}(v)$	Predictive entropy of v
$\mathbb{1}[\hat{y}_v \neq y_v]$	Misclassification indicator
$s(v)$	TAS-EGNN score for v
$\mathcal{S} \subset V$	Selected coreset

Consider an input graph $\mathcal{G} = (V, E, X, Y)$ with adjacency matrix A, where V is the set of n nodes, $E \subseteq V \times V$ is the edge set, $X \in \mathbb{R}^{n \times d}$ is the d-dimensional node feature matrix, and $Y \in \{1, \ldots, C\}^n$ is the label vector over C classes. Our goal is to identify a compact training subset $\mathcal{S} \subset V$ with $|\mathcal{S}| = c \ll n$

such that training a GNN on $\mathcal{S}$ achieves comparable performance to training on the full node set V. Unlike traditional coreset methods that select isolated nodes from global full-graph scores, TAS-EGNN operates on local *ego-graphs*, preserving contextual structural information during selection. We formulate coreset selection as a bilevel problem:

$$\begin{aligned} \min_{\mathcal{S}} \quad & \mathcal{L}_{\text{outer}}\big(f_{\theta_{\mathcal{S}}}(X, A),\, Y\big) \\ \text{s.t.} \quad & \theta_{\mathcal{S}} \in \arg\min_{\theta}\ \mathcal{L}_{\text{inner}}\big(f_{\theta}(X_{\mathcal{S}}, A_{\mathcal{S}}),\, Y_{\mathcal{S}}\big), \end{aligned} \tag{1}$$

where f_θ denotes a GNN and $\theta_{\mathcal{S}}$ are the parameters obtained by training only on the coreset $(X_{\mathcal{S}}, A_{\mathcal{S}}, Y_{\mathcal{S}})$. Unless otherwise stated, we chose both $\mathcal{L}_{\text{outer}}$, $\mathcal{L}_{\text{inner}}$, to be the cross-entropy loss ($\mathcal{L}_{\text{CE}}$). The formulation permits them to differ (e.g., a regularized or validation-style outer objective), but in our experiments, they are identical for a clean comparison. This optimization framework enables TAS-EGNN to produce subsets that retain both structural and task-relevant information, yielding scalable and effective GNN training under limited computational budgets.

2.1 Ego-Graphs in Spectral theory

Given the graph $\mathcal{G} = (V, E)$ with adjacency matrix A and degree matrix $D \in \mathbb{R}^{n\times n}$, $D_{ii} = \sum_j A_{ij}$, the normalized Laplacian is $L = I_n - D^{-1/2}AD^{-1/2}$, where I_n is the $n \times n$ identity matrix.

Ego-Graphs. Given a node $v \in V$ and hop radius r, the ego-graph $\mathcal{E}_v = (V_v, E_v)$ contains v, all its neighbors within r hops, and the edges among them. Let A_v and X_v denote the adjacency and feature matrices restricted to $\mathcal{E}_v$.

Spectral entropy on ego-graphs. Let L_v be the normalized Laplacian of $\mathcal{E}_v$, whose eigenvalues $\{\lambda_i\}_{i=1}^{k}$ lie in $[0, 2]$ for any undirected graph (the value 2 is attained iff the graph is bipartite). Interpreting the spectrum via normalized weights $\tilde{\lambda}_i = \lambda_i / \sum_{j=1}^{k} \lambda_j$, we define

$$\mathcal{H}_{\text{spec}}(\mathcal{E}_v) = -\sum_{i=1}^{k} \tilde{\lambda}_i \log \tilde{\lambda}_i \tag{2}$$

This treats the spectrum as a probability distribution over "graph frequencies", where $\mathcal{H}_{\text{spec}}$ is *high* when energy spreads across many Laplacian modes (e.g., around hubs, articulation points, or cross-community connectors), and *low* in regular, redundant regions such as near-cliques or uniform neighborhoods. High-entropy ego-graphs thus correspond to structurally heterogeneous areas that provide richer signal for message-passing, while low-entropy regions contribute largely redundant information. This interpretation aligns with prior work showing that eigenvalue-based Shannon entropy quantifies structural complexity and heterogeneity in graphs Passerini and Severini (2008); Dehmer and Mowshowitz (2011); De Domenico and Biamonte (2016). Moreover, our use of spectral entropy on r-hop ego-graphs (rather than the full graph Laplacian) reflects the localized nature of GNN message passing, where learning is driven primarily by neighborhood-level interactions. This intuition is consistent with works showing that local structural variation in ego-networks is highly predictive of model performance and embedding quality Pei et al. (2020); Zhu et al. (2021); Zeng et al. (2019). Finally, because eigenvalues are computed only on small ego-graphs (typically with k in the tens), the computation remains efficient ($O(k^3)$) and introduces negligible overhead relative to forward inference.

2.2 Predictive Uncertainty via Entropy

Selecting nodes based on task-specific difficulty is critical for maximizing learning utility. One key measure of task-related informativeness is *predictive uncertainty*, which reflects the confidence of the GNN in its predictions. Given a trained GNN model $f_\theta(A, X)$ that outputs softmax probabilities $p_v \in \mathbb{R}^C$ over C classes for each node v, the predictive uncertainty is quantified using Shannon entropy:

$$\mathcal{H}_{\text{pred}}(v) = -\sum_{c=1}^{C} p_v^{(c)} \log p_v^{(c)} \tag{3}$$

This entropy measures the information content of the predicted distribution: low entropy indicates confident predictions (i.e., highly skewed probability mass), whereas high entropy reflects uncertainty or confusion between classes. Nodes exhibiting high predictive uncertainty typically lie near class boundaries or in sparsely represented regions of the feature or structural space. Including such uncertain nodes in the coreset encourages the model to refine its classification boundaries, targeting regions where the current predictor is least confident. This principle is supported by classical results in active learning, where uncertainty sampling is shown to identify the most informative training samples for improving model decision boundaries Sener and Savarese (2018); Gal and Ghahramani (2016). By prioritizing high-uncertainty nodes, TAS-EGNN focuses training capacity on areas of maximal ambiguity, yielding more efficient and discriminative learning from a reduced sample set.

2.3 Supervised Signal from Misclassification Error

A third informative signal arises from actual model performance on labeled nodes. Nodes that are currently misclassified indicate areas where the model fails to generalize. To encode this error explicitly, we define a binary misclassification indicator:

$$\mathbb{1}[\hat{y}_v \neq y_v] = \begin{cases} 1 & \text{if } \hat{y}_v \neq y_v, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $\hat{y}_v = \arg\max_c p_v^{(c)}$ is the predicted label. This term directly injects supervised feedback into the coreset selection process, ensuring that nodes where the current model fails to generalize receive higher selection priority. Misclassified nodes typically correspond to hard or ambiguous cases that the model has not yet learned to distinguish, and training on such "hard examples" has been shown to significantly enhance classifier discrimination and convergence Bengio et al. (2009); Lin et al. (2017); Katharopoulos and Fleuret (2018). In this view, $\mathbb{1}[\hat{y}_v \neq y_v]$ acts as a binary hard-example indicator, enabling TAS-EGNN to explicitly focus training capacity on regions of high error, thereby accelerating learning and improving robustness.

2.4 Task-Aware Scoring Function

We now define the unified task-aware scoring function that incorporates the three complementary signals introduced above. For each node $v \in V_{\text{train}}$, its selection utility is quantified as follows by combining the three equations 2, 3, and 4:

$$s(v) = \alpha \cdot \mathcal{H}_{\text{spec}}(\mathcal{E}_v) + \beta \cdot \mathcal{H}_{\text{pred}}(v) + \gamma \cdot \mathbb{1}[\hat{y}_v \neq y_v], \tag{5}$$

where $\alpha, \beta, \gamma \geq 0$ are hyperparameters that control the importance of structural complexity, predictive uncertainty, and misclassification signal, respectively. The three components in this scoring function intentionally target different dimensions of node informativeness: structural distinctiveness through local spectral entropy ($\mathcal{H}_{\text{spec}}$), ambiguity near decision boundaries through predictive uncertainty ($\mathcal{H}_{\text{pred}}$), and hard examples through empirical error ($\mathbb{1}[\hat{y}_v \neq y_v]$). Their combination ensures that the selected coreset simultaneously captures structurally diverse regions of the graph, task-relevant areas of uncertainty, and challenging instances critical for improving model robustness and discrimination.

Algorithm 1 represents the steps of TAS-EGNN. This formulation enables a flexible and interpretable framework that balances both the structural and task-driven priorities in the coreset. TAS-EGNN uses GNN-specific task feedback (predictive uncertainty and misclassification) combined with local spectral complexity over ego-graphs to form a direct scoring function. Moreover, by selecting ego-graphs rather than isolated nodes, our approach captures richer local topologies, crucial for the expressive power of message-passing GNNs. This shift toward localized, supervision-aware scoring leads to coreset subsets that are not only structurally diverse but also finely tuned to the model's learning dynamics.

Algorithm 1 TAS-EGNN for Coreset Selection

Require: Graph $\mathcal{G} = (V, E, X, Y)$, GNN f_θ, scoring weights α, β, γ, coreset budget c
Ensure: Selected coreset $\mathcal{S}$
Initialize $\mathcal{S} \leftarrow \emptyset$
Compute predictions $\hat{Y} = f_\theta(X, A)$ and softmax probabilities p_v for all $v \in V_{\text{train}}$
for each node $v \in V_{\text{train}}$ **do**
  $\mathcal{E}_v \leftarrow r$-hop ego-graph
  $s(v) = \alpha \cdot \mathcal{H}_{\text{spec}}(\mathcal{E}_v) + \beta \cdot \mathcal{H}_{\text{pred}}(v) + \gamma \cdot \mathbb{1}[\hat{y}_v \neq y_v]$
end for
while $|\mathcal{S}| < c$ **do**
  $v^* \leftarrow \arg\max_{v \in V_{\text{train}} \setminus \mathcal{S}} \left[s(v) - \tau \sum_{u \in \mathcal{S}} \text{Overlap}(\mathcal{E}_u, \mathcal{E}_v) \right]$
  Update $\mathcal{S} \leftarrow \mathcal{S} \cup \{v^*\}$
end while
return $\mathcal{S}$

2.5 Greedy Algorithm for Coreset Optimization

Having defined the scoring function $s(v)$ for each candidate node, we aim to select a subset $\mathcal{S} \subset V_{\text{train}}$ that yields the most informative and structurally diverse training signal under a strict budget constraint. To this end, we formulate the coreset selection objective as:

$$\mathcal{S}^* = \arg\max_{\mathcal{S} \subseteq V_{\text{train}}} \left[\sum_{v \in \mathcal{S}} s(v) - \tau \sum_{(u,v) \in \mathcal{S}^2} \text{Overlap}(\mathcal{E}_u, \mathcal{E}_v) \right] \tag{6}$$

where $\text{Overlap}(\mathcal{E}_u, \mathcal{E}_v)$ measures shared nodes between the ego-graphs. The coefficient $\tau \geq 0$ modulates the influence of redundancy penalization relative to task-aware utility.

Greedy maximization and diminishing returns. Exact optimization of this objective is NP-hard, but the overlap-penalized marginal gain structure closely parallels well-studied monotone submodular maximization problems. In such settings, greedy selection is known to achieve a $(1 - 1/e)$ approximation to the optimal solution Nemhauser et al. (1978), and similar diminishing-return structures have been successfully employed in influence maximization Kempe et al. (2003), graph-based coverage Leskovec et al. (2007), and large-scale subset selection Mirzasoleiman et al. (2015). Motivated by this framework, we adopt a greedy selection strategy based on marginal informativeness:

$$v^* = \arg\max_{v \in V_{\text{train}} \setminus \mathcal{S}} \left[s(v) \; - \; \tau \sum_{u \in \mathcal{S}} \text{Overlap}(\mathcal{E}_u, \mathcal{E}_v) \right] \tag{7}$$

This iterative selection naturally encourages coverage of diverse structural regions while avoiding redundant

selections of highly overlapping ego-graphs. As a practical bookkeeping mechanism, we track the set of unique nodes contained in the selected ego-graphs, $\mathcal{U} = \bigcup_{v \in \mathcal{S}} \mathcal{E}_v$, and terminate once $|\mathcal{U}|$ reaches the target coreset size. This ensures that the final induced subgraph respects the node-level training budget, enabling valid comparisons against coreset, condensation, and coarsening approaches while maintaining strong task relevance and structural diversity.

2.6 Complexity Analysis

Let d denote the feature dimension, L the GNN layers, V_{train} the candidate set, and $n{=}|V|$, $m{=}|E|$. For each node v, let $k_v{:=}|V_v|$ be the size of its r-hop ego-graph and $\bar{k}$ the average k_v (typically tens). TAS-EGNN performs (i) one backbone forward pass to obtain p_v and $\hat{y}_v$ for all $v \in V_{\text{train}}$, costing $O(L\,m\,d)$; (ii) spectral-entropy computation on each ego-graph, $O(k_v^3)$ per v (exact eigenvalues on small k_v), totaling $\sum_{v \in V_{\text{train}}} O(k_v^3)$; and (iii) a greedy coverage step with a union-size budget U, where evaluating each marginal gain using a maintained covered set takes $O(k_v)$ per candidate, giving $O(c'\,|V_{\text{train}}|\,\bar{k})$ over c' selections (with $c' \approx U/\bar{k}$ in practice). Overall,

$$T_{\text{total}} \;=\; O(L\,d\,m) \;+\; \sum_{v \in V_{\text{train}}} O(k_v^3) \;+\; O\big(c'\,|V_{\text{train}}|\,\bar{k}\big),$$

which is near-linear in m (single forward pass) and linear in the effective budget U for fixed small r. Since the training process occurs on the induced union of selected ego-graphs, the memory complexity is $O(n\,d + m) + O(U)$.

3 EVALUATION AND VALIDATION

We conduct comprehensive experiments to evaluate the effectiveness of TAS-EGNN in selecting task-aware and structurally diverse coresets for graph neural network training. Our evaluation spans standard benchmarks, aiming to demonstrate both efficiency and predictive performance under various coreset budgets. We compare TAS-EGNN against three families of coreset-related methods: (i) random and distribution-based sampling (Random, Herding Welling (2009), K-Center Farahani and Hekmatfar (2009); Sener and Savarese (2017)), which ignore graph structure and treat nodes as i.i.d.; (ii) graph condensation approaches (GCond Jin et al. (2021), CGC Gao et al. (2025)), which attempt to generate synthetic or prototype nodes; and (iii) graph coarsening selection methods (SGBGC Xia et al. (2025), UGC Kataria et al. (2024)), and ego-graph corsets method SGGC Ding et al. (2024), which perform structure-aware selection but do not incorporate task-driven predictive feedback. This positioning highlights the unique role of TAS-EGNN as a method that jointly exploits structural heterogeneity and model-specific uncertainty. Further details about the baseline techniques, datasets, and implementation details used in our experiments are provided in the App. B. As an evaluation of the performance of TAS-EGNN, first, we provide a comparison between the performance of two GNN backbones under TAS-EGNN, which are GCN and GraphSAGE, in Sec. 3.1. Additional experimental results using a larger set of GNNs (APPNP, GAT, GCN, GIN, GraphSAGE) are given in the App. B.1. Second, an ablation of the scoring components (Sec. 3.2), then we report node-classification results on citation, social, products, and transaction-fraud detection graphs (Sec. 3.3), and finally, a comprehensive efficiency profile in time and memory is presented in Sec. 3.4. Furthermore, we provide the Hyperparameter Sensitivity evaluation and analysis in the App. B.2.

3.1 Backbone-Agnostic Integration of TAS-EGNN

A defining property of TAS-EGNN is that the coreset selection mechanism is fully decoupled from the downstream GNN predictor: it operates entirely on ego-graph structure, predictive entropy, and misclassification signals, without assuming any particular neural architecture. As a result, the selected coreset can be consumed by *any* message-passing GNN. To validate this backbone independence, we applied TAS-EGNN using two representative architectures, GCN Kipf and Welling (2016) and GraphSAGE Hamilton et al. (2017), across all datasets and coreset ratios. The results, shown in Tab. 2, demonstrate that TAS-EGNN delivers consistent performance gains across both backbones and domains. Notably, while GraphSAGE often yields the strongest results, particularly on large-scale or structurally heterogeneous graphs, the improvement observed with GCN confirms that TAS-EGNN transfers its benefits regardless of the underlying architecture. This supports our claim that TAS-EGNN is inherently backbone-agnostic and can be paired with various message-passing models. A more extensive backbone study, including GAT, APPNP, and GIN, is provided in App. B.1.

3.2 Ablation Study

Settings. We probe the contribution of each scoring component in TAS-EGNN by ablating the ego-graph *spectral entropy* $\mathcal{H}_{\text{spec}}$, *predictive uncertainty* $\mathcal{H}_{\text{pred}}$, and the *misclassification* signal $\mathbb{1}[\hat{y} \neq y]$. Concretely, we compare the full score $s(v) = \alpha\,\mathcal{H}_{\text{spec}}(\mathcal{E}_v) + \beta\,\mathcal{H}_{\text{pred}}(v) + \gamma\,\mathbb{1}[\hat{y}_v \neq y_v]$ against its two-term and single-term variants, holding the backbone, optimizer, coreset budget, and greedy coverage procedure fixed across conditions. Following standard practice, we report *Accuracy* at $r = 50\%$ on Cora and Citeseer, $r = 2\%$ on OGBN-Arxiv and Flickr, $r = 0.2\%$ on OGBN-Products, and $r = 50\%$ on Banksim, Paysim, and ECC datasets. The results are mean $\pm$ std over repeated runs (Tab. 3).

Table 2: Backbone comparison of TAS-EGNN using GCN and GraphSAGE across all datasets and coreset ratios. Benchmark Graph datasets are evaluated with Accuracy, while Transaction-Fraud datasets are evaluated with PR-AUC. Best results for each dataset/ratio are in bold.

Dataset	Ratio	GCN	GraphSAGE
Citeseer	15%	57.2±0.5	**65.4±1.2**
	25%	71.5±0.3	**71.7±0.9**
	50%	77.1±0.4	**80.3±0.5**
Cora	15%	69.4±0.2	**83.9±0.1**
	25%	82.8±0.1	**87.9±0.1**
	50%	86.8±0.1	**92.3±0.2**
Flickr	0.5%	56.8±0.3	**72.9±1.2**
	1.0%	58.9±0.3	**74.4±1.7**
	2.0%	59.7±0.2	**75.8±1.6**
OGBN-Arxiv	0.5%	62.5±0.2	**66.6±0.2**
	1.0%	64.3±0.1	**67.4±0.2**
	2.0%	65.1±0.2	**68.6±0.1**
OGBN-Products	0.1%	61.1±0.3	**64.8±0.5**
	0.2%	65.2±0.4	**69.5±0.6**
Banksim	15%	92.8±3.3	**96.3±0.2**
	25%	93.0±2.5	**96.4±0.2**
	50%	93.1±1.5	**96.5±0.5**
Paysim	15%	94.0±1.6	**94.6±1.1**
	25%	95.4±2.6	**96.0±1.0**
	50%	95.6±0.9	**96.1±1.4**
ECC	15%	91.8±0.7	**97.4±0.5**
	25%	92.4±0.9	**97.5±0.6**
	50%	93.8±0.6	**97.6±0.2**

Results and Discussion. Across all eight datasets, the full TAS-EGNN score $s(v)$ attains the best or tied-best result, confirming that the three cues are complementary. Among single terms, *predictive uncertainty* is generally the strongest: on *Citeseer*, *Cora*, *Flickr*, *OGBN-Arxiv*, and *OGBN-Products* it reaches 79.1%, 91.7%, 75.4%, 68.3%, and 69.2% respectively, surpassing either spectral-only or error-only. On transaction-fraud graphs, uncertainty alone is competitive on *Paysim* (95.7%) and *ECC* (95.1%), though on *Banksim* structural information is more decisive (spectral-only 95.1% vs. pred-only 93.5%). Pairwise combinations narrow the gap to the full score: *Spec+Pred* is within 0.1–0.2 points on *Citeseer* and *Cora*, and reaches 96.0% on *Banksim* and 97.2% on *ECC*. The *error* term is weakest in isolation (e.g., 75.0% on *Citeseer*, 88.6% on *Cora*), but it contributes small, consistent gains when added to the other cues: on *Flickr* it lifts *Spec+Pred* from 74.8% to 75.8%, and on *ECC* from 97.2% to 97.6%. On *OGBN-Arxiv* and *OGBN-Products*, all variants are tightly clustered (68.0–68.6% and 69.2–69.5% respectively), indicating robustness at this operating point while the full score remains highest. As a clear pattern, the uncertainty provides the most reliable single signal across domains; spectral entropy becomes increasingly valuable on large, sparse, or structurally irregular graphs; and the misclassification indicator, though noisy alone, adds complementary task feedback that nudges performance to the top when all three are combined.

Importantly, these ablation patterns reveal how the different components of the TAS-EGNN score interact in a complementary manner rather than redundantly. Nodes prioritized by $\mathcal{H}_{\text{spec}}$ tend to arise from topologically heterogeneous regions, but are not necessarily uncertain; nodes selected by $\mathcal{H}_{\text{pred}}$ frequently lie near decision boundaries, but are not structurally unique; and nodes identified by $\mathbb{1}[\hat{y}_v \neq y_v]$ explicitly highlight regions where the current model fails. The full score effectively aggregates these signals, yielding a coreset that is simultaneously structurally diverse, uncertainty-aware, and error-focused, leading to the consistently superior performance observed in Table 3.

3.3 Node Classification Performance

This section reports node classification results on the benchmark families and aligns the evaluation protocol to each. (i) *Benchmark Graph Datasets*: Cora, Citeseer, Flickr, OGBN-Arxiv, and OGBN-Products, where we measure accuracy (ACC). Following prior work, we use budgets $r \in \{15\%, 25\%, 50\%\}$ on Cora and Citeseer, $r \in \{0.5\%, 1.0\%, 2.0\%\}$ on OGBN-Arxiv and Flickr, and $r \in \{0.1\%, 0.2\%\}$ on OGBN-Products to reflect their scale. (ii) *Transaction–fraud graphs*, Banksim, Paysim, and ECC, constructed as account–transaction networks with feature aggregates (amount, count, type, recency), where anomalies are fraudulent entities. Due to severe class imbalance, we report PR-AUC as the primary metric. Across both families, we compare TAS-EGNN to classical coresets (Random, Herding, K-Center), graph condensation (GCond, CGC), graph coarsening (SGBGC, UGC), and ego-graph coresets (SGGC). For each method and budget, we select a coreset from the training split and train the *same* downstream classifier with fixed hyperparameters. The results are presented first for **Benchmark Graph Datasets** (Tab. 4), then for **Transaction–Fraud Detection datasets** (Tab. 5).

Benchmark Graph Datasets. Table 4 presents the accuracy of TAS-EGNN and all baselines across five datasets and multiple coreset ratios. Several insights emerge. First, across all datasets and ratios, classical techniques, Random, Herding, and K-Center, consistently underperform. Their ignorance of graph structure and task dynamics limits their capacity to extract informative coresets, especially in low-budget settings (e.g., 15% or 0.5%). Graph condensation methods, particularly CGC, generally outperform classical baselines due to their label-aware condensation and class-prototype modeling. GCond performs well on smaller graphs but fails to scale on OGBN-Arxiv, yielding out-of-memory (OOM) errors at higher ratios, highlighting a key limitation in gradient-based condensation pipelines. Among graph coarsening methods, UGC shows strong performance on citation graphs like Citeseer and Cora, especially at low to mid-range budgets, reflecting its spectral preservation strength.

Table 3: Ablation on TAS-EGNN score components in terms of accuracy metric using ratios $r = 50\%$ for Cora and Citeseer, $r = 2.0\%$ for OGBN-Arxiv and Flickr datasets, and $r = 0.2\%$ for OGBN-Products, and PR-AUC metric on Transaction-Fraud Detection datasets using ratio $r = 50\%$. Best in bold.

Variant	Citeseer	Cora	Flickr	OGBN-Arxiv	OGBN-Products	Banksim	Paysim	ECC
$s(v)$	**80.3±0.5**	**92.3±0.2**	**75.8±1.6**	**68.6±0.1**	**69.5±0.6**	**96.5±0.5**	**96.1±1.4**	**97.6±0.2**
$\alpha \cdot \mathcal{H}_{\text{spec}}(\mathcal{E}_v) + \beta \cdot \mathcal{H}_{\text{pred}}(v)$	80.2±0.4	92.2±0.1	74.8±0.8	68.5±0.1	69.4±0.5	96.0±0.3	96.0±1.2	97.2±0.5
$\alpha \cdot \mathcal{H}_{\text{spec}}(\mathcal{E}_v) + \gamma \cdot \mathbb{1}[\hat{y}_v \neq y_v]$	79.0±0.3	90.4±0.1	75.3±1.4	68.5±0.1	69.3±0.7	95.6±0.4	95.8±1.5	97.4±0.3
$\beta \cdot \mathcal{H}_{\text{pred}}(v) + \gamma \cdot \mathbb{1}[\hat{y}_v \neq y_v]$	80.1±0.2	91.9±0.2	75.5±0.6	68.5±0.2	69.4±0.2	93.7±1.1	96.0±1.3	97.0±0.3
$\mathcal{H}_{\text{spec}}(\mathcal{E}_v)$	78.9±0.2	90.3±0.1	74.1±1.8	68.5±0.3	69.3±0.1	95.1±1.3	95.1±1.5	94.4±2.4
$\mathcal{H}_{\text{pred}}(v)$	79.1±0.1	91.7±0.2	75.4±0.7	68.3±0.1	69.2±0.5	93.5±0.9	95.7±0.6	95.1±0.5
$\mathbb{1}[\hat{y}_v \neq y_v]$	75.0±0.4	88.6±0.1	75.0±1.1	68.0±0.1	69.2±0.4	93.4±0.7	95.7±1.1	94.2±1.6

Table 4: Performance classification comparison of TAS-EGNN against the baselines in terms of the accuracy metric. Best in bold. Δ is the gap between TAS-EGNN and the best competitor performance.

		Classical Coresets			Graph Condensation		Graph Coarsening		Ego-Graph Coresets		
Dataset	Ratio	Random	Herding	K-Center	GCOND	CGC	SGBGC	UGC	SGGC	TAS-EGNN	Δ
Citeseer	15%	41.9±1.1	46.1±1.6	47.5±6.3	71.9±1.3	71.6±0.3	74.3±1.1	**76.2±0.1**	63.7±3.1	65.4±1.2	-10.8
	25%	54.4±4.4	54.9±3.9	61.6±4.0	70.5±1.2	72.5±0.2	75.5±0.8	**77.1±0.1**	67.2±2.4	71.7±0.9	-5.4
	50%	64.2±1.7	68.7±0.5	65.6±1.6	70.6±0.9	72.4±0.2	75.3±0.9	70.3±0.0	68.4±0.9	**80.3±0.5**	+5.0
Cora	15%	54.0±1.6	66.1±1.2	64.3±4.8	79.8±0.6	82.7±0.3	83.8±0.9	**84.7±0.2**	72.9±0.6	83.9±0.1	-0.8
	25%	63.6±3.7	69.9±1.0	72.6±2.5	79.8±1.3	82.3±1.3	86.1±1.0	84.9±0.0	78.6±1.0	**87.9±0.1**	+1.8
	50%	72.8±1.1	70.8±0.4	78.9±1.1	80.1±0.6	82.5±0.6	87.7±0.9	83.1±0.1	80.2±0.8	**92.3±0.1**	+4.6
Flickr	0.5%	43.1±0.3	45.5±1.3	46.9±0.9	46.5±0.4	46.8±0.0	42.3±0.0	OOM	48.4±0.8	**72.9±1.2**	+24.5
	1.0%	42.7±0.3	46.7±0.3	47.5±0.9	47.1±0.1	47.1±0.1	42.3±0.0	OOM	49.0±0.6	**74.4±1.7**	+25.4
	2.0%	43.6±0.5	45.5±0.6	46.9±0.7	45.4±0.0	47.0±0.1	43.6±0.2	OOM	64.4±0.4	**75.8±1.6**	+11.4
OGBN-Arxiv	0.5%	48.3±0.5	45.7±4.4	56.8±2.8	63.2±0.3	66.4±0.1	28.7±1.2	OOM	59.7±1.5	**66.6±0.2**	+0.2
	1.0%	51.4±0.4	47.6±0.4	60.7±0.8	64.0±0.4	67.2±0.4	31.1±1.8	OOM	62.5±0.9	**67.4±0.2**	+0.2
	2.0%	54.6±0.4	56.5±0.5	62.4±0.9	OOM	67.6±0.2	34.3±1.2	OOM	64.4±0.4	**68.6±0.1**	+1.0
OGBN-Products	0.1%	36.8±0.4	38.3±0.3	35.2±0.8	OOM	OOM	OOM	OOM	OOM	**64.8±0.5**	+26.5
	0.2%	41.8±0.8	45.0±0.4	38.2±0.4	OOM	OOM	OOM	OOM	OOM	**69.5±0.6**	+24.5

Table 5: Performance classification comparison of TAS-EGNN against the baselines in terms of the PR-AUC metric. Best in bold. Δ is the gap between TAS-EGNN and the best competitor performance.

		Classical Coresets			Graph Comdensation		Graph Coarsening		Ego-Graph Coresets		
Dataset	Ratio	Random	Herding	K-Center	GCOND	CGC	SGBGC	UGC	SGGC	TAS-EGNN	Δ
Banksim	15%	86.5±0.3	91.5±0.2	90.3±0.1	84.5±1.4	93.1±0.1	93.1±0.1	94.8±0.1	95.7±0.1	**96.3±0.2**	0.6
	25%	86.6±0.1	92.8±0.1	90.5±0.1	OOM	93.6±0.1	92.9±0.1	94.8±0.1	95.8±0.1	**96.4±0.2**	0.6
	50%	86.7±0.1	93.2±0.1	90.6±0.1	OOM	93.5±0.1	92.8±0.1	94.2±0.1	93.5±0.2	**96.5±0.5**	2.3
Paysim	15%	79.2±0.1	91.0±0.3	90.4±0.0	84.3±0.5	74.0±0.4	94.3±0.1	93.6±0.0	48.4±0.3	**94.6±1.1**	0.3
	25%	79.4±0.1	91.2±0.2	90.5±0.1	OOM	75.4±0.7	93.2±0.2	93.5±0.0	45.2±0.1	**96.0±1.0**	2.5
	50%	79.5±0.1	92.4±0.1	90.5±0.1	OOM	79.0±0.5	92.3±0.2	93.3±0.0	45.0±0.2	**96.1±1.4**	2.8
ECC	15%	80.7±0.2	90.6±0.3	89.5±0.1	OOM	86.9±0.3	82.5±0.5	79.4±0.6	50.0±0.3	**97.4±0.5**	6.8
	25%	81.1±0.1	90.1±0.5	90.2±0.2	OOM	83.4±1.5	83.2±0.5	79.0±1.2	52.0±0.2	**97.5±0.6**	7.3
	50%	81.9±0.1	91.6±0.2	90.5±0.1	OOM	84.4±0.6	78.7±0.3	79.0±1.3	52.6±0.2	**97.6±0.2**	6.0

However, UGC fails to scale to large graphs such as OGBN-Arxiv and Flickr. SGBGC performs reasonably well on Citeseer and Cora but struggles on larger graphs, especially OGBN-Arxiv, where its accuracy drops sharply. SGGC, the most direct competitor in ego-graph coreset selection, performs well on Cora and Citeseer but shows noticeable degradation on OGBN-Arxiv and Flickr. This suggests that while spectral alignment is effective in small graphs, it may not fully capture task-relevant substructures at scale. On the *OGBN-Products* dataset, nearly all competing methods fail to execute due to memory limitations, including graph condensation (GCond, CGC), graph coarsening (UGC, SGBGC), and even prior ego-graph selection (SGGC). In contrast, TAS-EGNN consistently achieves the best or near-best performance across all datasets and coreset ratios. On Cora and Citeseer, it surpasses all baselines at 50% budget, achieving 92.3% and 80.3% accuracy, respectively. On OGBN-Arxiv, it delivers 68.6% accuracy at 2.0%, outperforming CGC and avoiding the scalability bottlenecks of UGC and GCond. Notably, on Flickr, TAS-EGNN achieves 75.8% accuracy at 2.0%, a significant margin above all baselines, including SGGC, whose spectral focus limits task-specific generalization in sparse, large-scale networks. Regarding *OGBN-Products*, TAS-EGNN successfully operates at extremely small budgets (0.1% and 0.2%), achieving 64.8% and 69.5% accuracy, respectively, while classical selection strategies remain below 45%. The column Δ quantifies the performance gap between the proposed algorithm and the best competitor. *Citeseer*, it grows from -10.8 (at 15%) and -5.4 (at 25%) to 5.0 (at 50%); on *Cora* it increases $-0.8 \rightarrow 1.8 \rightarrow 4.6$ as the budget rises, indicating widening margins at higher ratios. On

OGBN-Arxiv, Δ is small (0.2, 0.2, 1.0), reflecting a tight top tier where TAS-EGNN still leads. On *Flickr* and *OGBN-Products*, Δ is very large with values 11.4-25.4 for Flickr and 24.5-26.5 for OGBN-Products, showing decisive gains in the low-budget, large sparse regime. The Δ gap confirms that TAS-EGNN's task-aware scoring and greedy coverage yield increasingly pronounced advantages on citation graphs as the budget grows and deliver clear separation on large social graphs even at tiny budgets. These results show the ability of TAS-EGNN to balance structural diversity, predictive uncertainty, and misclassification awareness, yielding coresets that retain rich learning signals even under tight budget constraints. Its principled scoring strategy and efficient greedy coverage mechanism make it well-suited for real-world scenarios involving massive, heterogeneous graphs.

Transaction-Fraud Detection Graphs. Table 5 shows that **TAS-EGNN** attains the best PR-AUC across all fraud datasets and budgets. On *Banksim*, TAS-EGNN yields 96.3, 96.4, and 96.5 at 15%, 25%, and 50%, respectively, outperforming the closest baseline (SGGC or UGC depending on the budget) by $\approx$ 0.6 to 2.3 points; correspondingly, the gap column reports $\Delta = 0.6$, 0.6, and 2.3, indicating a small but consistent lead at lower budgets that widens at 50%. On *Paysim*, gains widen with budget (94.6 $\rightarrow$ 96.1), surpassing the strongest baseline (SGBGC/UGC) by $\approx$ 0.3 to 2.8; here Δ grows from 0.3 to 2.5 and 2.8, showing increasing separation as more budget is available. The largest margins appear on *ECC*, where TAS-EGNN reaches 97.4–97.6, exceeding the best non–ego-graph method (e.g., Herding $leftarrow$ 90–92) by $\rightarrow$ 6–7.4 points; the Δ values of 6.8, 7.3, and 6.0 confirm a stable, substantial advantage across all budgets. Classical coresets (Random, Herding, K-Center) improve with budget but plateau. Graph condensation (GCOND) frequently runs out of memory at 25%–50% on our hardware, and coarsening methods (UGC, SGBGC) are dataset-dependent, UGC is competitive on *Banksim* but substantially underperforms on *ECC*. SGGC degrades on *Paysim* and *ECC*. The consistency of TAS-EGNN across three distinct transaction graphs, together with the positive Δ gaps (especially the strong margins on *ECC*), suggests that selecting diverse, task-informative ego-regions is particularly effective for imbalanced fraud detection.

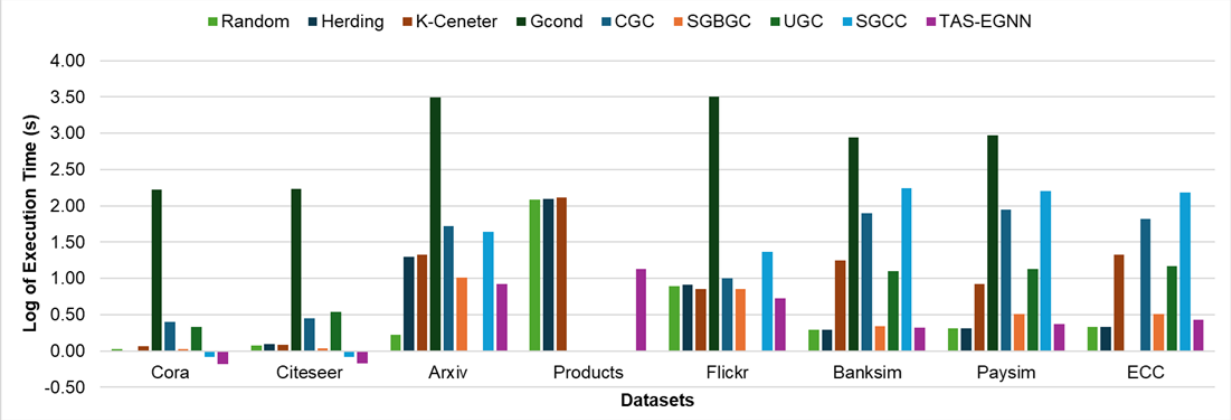
3.4 Computational Efficiency and Memory Usage

We benchmark end-to-end efficiency for two benchmark families using a common machine and fixed training protocol. For each dataset, we measure (i) the graph-reduction step at r=50% on Cora and Citeseer, r=1.0% on OGBN-Arxiv and Flickr, r=0.2% on OGBN-Products, and r=15% on Banksim, Paysim, and ECC, plus (ii) we train the same downstream model on the reduced set. We report $\log_{10}$ of time (seconds) in Fig. 1a and $\log_{10}$ of GPU memory (MB) in Fig. 1b. If a method exceeds a time or memory cap, we mark it OOT/OOM and clip at the axis limit. Figure 1 represents the results of TAS-EGNN against the baselines on all benchmark datasets.

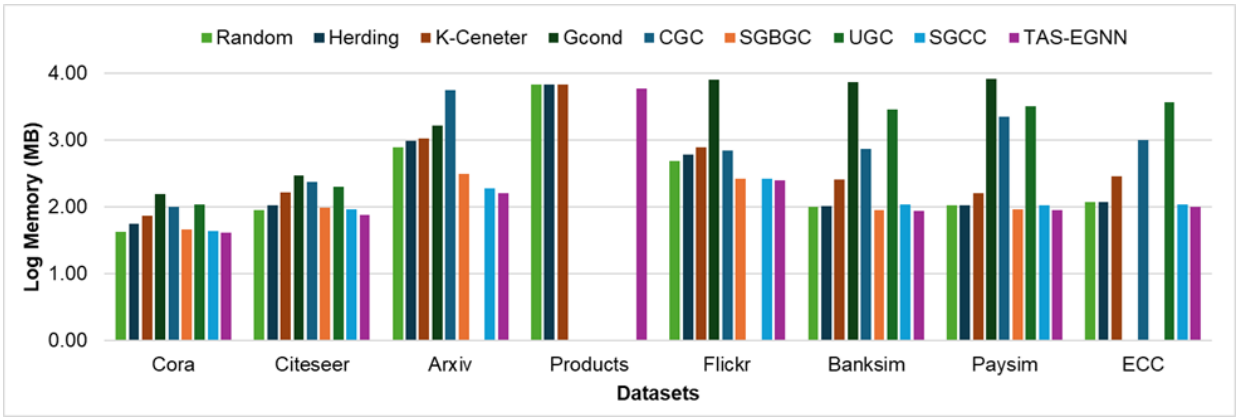
Benchmark Graph Datasets. Figure 1a represents the log_{10} of execution time on all datasets. The performance of all methods is fast on Cora and Citeseer (small spread on the log scale). As the graph size grows, as is the case for OGBN-Arxiv and Flickr datasets, the differences widen sharply. For the *OGBN-Products* dataset, the runtime comparison is restricted to TAS-EGNN and the classical coreset selection methods (Random, Herding, and K-Center) since the other approaches could not be executed on this dataset due to memory and computational constraints on available hardware. TAS-EGNN is substantially faster while simultaneously achieving significantly higher accuracy compared to TAS-EGNN. GCond incurs the largest runtime growth, reflecting the cost of gradient-matching condensation. UGC escalates dramatically, hitting the OOT cap on large graphs, consistent with its heavier spectral objectives. SGBGC, CGC, and SGGC rise moderately. TAS-EGNN stays on, or very near, the lower envelope across datasets: it is close to cheap classical baselines (Random, Herding, and K-Center) on the small graphs, and remains substantially faster than condensation and coarsening methods on OGBN-Arxiv and Flickr datasets. This indicates that ego-graph scoring plus greedy coverage yields strong task awareness with minimal overhead.

Figure 1b shows the results of the TAS-EGNN against the baseline techniques in terms of memory usage. The memory landscape mirrors time. GCond shows the steepest VRAM growth and approaches the top of the axis on Flickr, while UGC spikes early and becomes impractical on large graphs. SGBGC, CGC, and SGGC increase with graph size but remain below GCond and UGC. TAS-EGNN maintains one of the lowest memory footprints among structure-aware methods across all datasets; only the trivial classical coresets, e.g., Random, Herding, and K-Center, which ignore graph structure and task signals, are consistently lighter. This aligns with TAS-EGNN's localized computations on small ego-graphs instead of global eigensolvers or bilevel optimization.

Transaction-Faud Detection Graphs. Figures 1a and 1b represent the time and memory ($\log_{10}$ scale) reinforces the scalability picture. GCOND exhibits the steepest growth and hits OOM in several settings, aligning with its heavy bilevel and gradient pipeline. UGC shows early spikes in both runtime and VRAM, becom-



(a) Log_{10} of execution time of all datasets.



(b) Log_{10} of memory usage of all datasets.

Figure 1: Runtime and Memory ($\log_{10}$ scale) efficiency of TAS-EGNN against all baseline techniques using ratios $r = 50\%$ for Cora and Citeseer, $r = 1.0\%$ for OGBN-Arxiv and Flickr, $r = 0.2\%$ for OGBN-Products, $r = 15\%$ for Banksim, Paysim, and ECC datasets (see online for colors).

ing impractical on larger graphs. SGBGC, CGC, and SGGC grow with graph size but remain below GCOND and UGC. TAS-EGNN maintains one of the lowest footprints among structure-aware methods across datasets, trailing only the trivial classical coresets, e. i, Random, Herding, and K-Center, which do not model graph structure or task signals. This efficiency stems from TAS-EGNN's localized computations on the ego-graphs and a single forward pass for uncertainty and error signals, avoiding global eigensolvers or bilevel optimization. TAS-EGNN, on imbalanced transaction graphs, delivers state-of-the-art PR-AUC while preserving low time and memory scaling like classical coresets, yet retaining the structural and task awareness needed for fraud detection.

4 CONCLUSION

We introduced TAS-EGNN, a task-aware, ego-graph–based coreset selection framework that unifies local spectral complexity, predictive uncertainty, and misclassification feedback into a single scoring function, followed by a greedy coverage step for non-redundant selection. By operating on lightweight ego-graphs and avoiding global eigendecomposition or heavy bilevel optimization, TAS-EGNN is both principled and efficient. Empirically, TAS-EGNN delivers strong performance across different benchmark families. On citation, social, and products graphs (Cora, Citeseer, Flickr, OGBN-Arxiv, and OGBN-Products), it achieves state-of-the-art or near-state-of-the-art accuracy under tight budgets. On transaction–fraud detection graphs (Banksim, Paysim, ECC), it attains leading PR-AUC under severe class imbalance while preserving a low computational footprint. Our profiling shows that TAS-EGNN tracks the lower envelope among structure-aware methods in *log*-scaled execution time and peak GPU memory, and scales to large graphs where several baselines reach OOT/OOM. The method is robust across GNN backbones, with GraphSAGE particularly effective in large inductive settings. TAS-EGNN advances scalable, task-aware coreset selection, yielding practical accuracy–efficiency trade-offs for real-world graph learning. Future directions include (i) dynamic streaming extensions with amortized updates of ego-graph scores, (ii) tighter theory for coverage and utility guarantees and faster moment-based spectral proxies, and (iii) broader high-stakes applications (beyond fraud) where PR-AUC and resource constraints are critical.

Acknowledgements

This work was supported by a research grant (No. IT43071) funded by Mitacs and Mastercard.

References

Hajar Saif Alsaadi, Rachid Hedjam, Abderezak Touzene, and Abdelhamid Abdessalem. Fast binary network intrusion detection based on matched filter optimization. In *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, pages 195–199. IEEE, 2020.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

Andrea Dal Pozzolo, Olivier Caelen, Yann-Ael Le Borgne, Serge Waterschoot, and Gianluca Bontempi. Learned lessons in credit card fraud detection from a practitioner perspective. *Expert systems with applications*, 41:4915–4928, 2014.

Manlio De Domenico and Jacob Biamonte. Spectral entropies as information-theoretic tools for complex network comparison. *Physical Review X*, 6:041062, 2016.

Matthias Dehmer and Abbe Mowshowitz. A history of graph entropy measures. *Information Sciences*, 181: 57–78, 2011.

Mucong Ding, Yinhan He, Jundong Li, and Furong Huang. Spectral greedy coresets for graph neural networks. *arXiv preprint arXiv:2405.17404*, 2024.

Keyu Duan, Zirui Liu, Peihao Wang, Wenqing Zheng, Kaixiong Zhou, Tianlong Chen, Xia Hu, and Zhangyang Wang. A comprehensive study on large-scale graph training: Benchmarking and rethinking. *Advances in Neural Information Processing Systems*, 35:5376–5389, 2022.

Abhimanyu Dubey, Moitreya Chatterjee, and Narendra Ahuja. Coreset-based neural network compression. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 454–470, 2018.

Reza Zanjirani Farahani and Masoud Hekmatfar. *Facility location: concepts, models, algorithms and case studies*. Springer Science & Business Media, 2009.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

Xinyi Gao, Guanhua Ye, Tong Chen, Wentao Zhang, Junliang Yu, and Hongzhi Yin. Rethinking and accelerating graph condensation: A training-free approach with class partition. In *Proceedings of the ACM on Web Conference 2025*, pages 4359–4373, 2025.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv preprint arXiv:1810.05997*, 2018.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

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

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, 33:22118–22133, 2020.

Wei Jin, Lingxiao Zhao, Shichang Zhang, Yozen Liu, Jiliang Tang, and Neil Shah. Graph condensation for graph neural networks. *arXiv preprint arXiv:2110.07580*, 2021.

Yu Jin, Andreas Loukas, and Joseph JaJa. Graph coarsening with preserved spectral properties. In *International Conference on Artificial Intelligence and Statistics*, pages 4452–4462. PMLR, 2020.

Antonin Joly and Nicolas Keriven. Graph coarsening with message-passing guarantees. *Advances in Neural Information Processing Systems*, 37:114902–114927, 2024.

Mohit Kataria, Sandeep Kumar, et al. Ugc: Universal graph coarsening. *Advances in Neural Information Processing Systems*, 37:63057–63081, 2024.

Angelos Katharopoulos and François Fleuret. Not all samples are created equal: Deep learning with importance sampling. In *International conference on machine learning*, pages 2525–2534. PMLR, 2018.

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

Krishnateja Killamsetty, Xujiang Zhao, Feng Chen, and Rishabh Iyer. Retrieve: Coreset selection for efficient and robust semi-supervised learning. *Advances in neural information processing systems*, 34: 14488–14501, 2021.

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

Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429, 2007.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

Edgar Lopez-Rojas, Ahmad Elmir, and Stefan Axelsson. Paysim: A financial mobile money simulator for fraud detection. In *28th European Modeling and Simulation Symposium, EMSS, Larnaca*, pages 249–255. Dime University of Genoa, 2016.

Edgar Alonso Lopez-Rojas and Stefan Axelsson. Banksim: A bank payments simulator for fraud detection research. In *26th European Modeling and Simulation Symposium*, page 144–152, 2014.

Baharan Mirzasoleiman, Ashwinkumar Badanidiyuru, Amin Karbasi, Jan Vondrák, and Andreas Krause. Lazier than lazy greedy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

Baharan Mirzasoleiman, Jeff Bilmes, and Jure Leskovec. Coresets for data-efficient training of machine learning models. In *International Conference on Machine Learning*, pages 6950–6960. PMLR, 2020.

Manish Nagaraj, Deepak Ravikumar, Efstathia Soufleri, and Kaushik Roy. Finding the muses: Identifying coresets through loss trajectories. *arXiv preprint arXiv:2503.09721*, 2025.

Saket Navlakha, Rajeev Rastogi, and Nisheeth Shrivastava. Graph summarization with bounded error. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 419–432, 2008.

George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming*, 14:265–294, 1978.

Filippo Passerini and Simone Severini. The von neumann entropy of networks. *arXiv preprint arXiv:0812.2597*, 2008.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. *ICLR*, 2020.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *ICLR*, 2018.

Ruilin Tong, Yuhang Liu, Javen Qinfeng Shi, and Dong Gong. Coreset selection via reducible loss in continual learning. In *The Thirteenth International Conference on Learning Representations*, 2025.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.

Lilapati Waikhom and Ripon Patgiri. A survey of graph neural networks in various learning paradigms: methods, applications, and challenges. *Artificial Intelligence Review*, 56:6295–6364, 2023.

Max Welling. Herding dynamical weights to learn. In *Proceedings of the 26th annual international conference on machine learning*, pages 1121–1128, 2009.

Shuyin Xia, Xinjun Ma, Zhiyuan Liu, Cheng Liu, Sen Zhao, and Guoyin Wang. Graph coarsening via supervised granular-ball for scalable graph neural network training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 12872–12880, 2025.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations (ICLR)*, 2018.

Hanqing Zeng, Hongkuan Zhou, Ajitesh Srivastava, Rajgopal Kannan, and Viktor Prasanna. Graphsaint: Graph sampling based inductive learning method. *arXiv preprint arXiv:1907.04931*, 2019.

Hanqing Zeng, Hongkuan Zhou, Ajitesh Srivastava, Rajgopal Kannan, and Viktor Prasanna. Graphsaint: Graph sampling based inductive learning method. In *International Conference on Learning Representations (ICLR)*, 2020.

Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 793–803, 2019.

Qi Zhu, Carl Yang, Yidan Xu, Haonan Wang, Chao Zhang, and Jiawei Han. Transfer learning of graph neural networks with ego-graph information maximization. *Advances in Neural Information Processing Systems*, 34:1766–1779, 2021.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Not Applicable]
   (b) Complete proofs of all theoretical results. [Not Applicable]
   (c) Clear explanations of any assumptions. [Not Applicable]
3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]
5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

TAS-EGNN: Task-Aware Spectral Ego-Graphs for Efficient GNNs-Based Classification Supplementary Materials

A Related Work

Coreset selection and graph reduction strategies have emerged as vital approaches for addressing the scalability challenges of Graph Neural Networks (GNNs). In this study, we categorize existing methods into four main paradigms: Classical Coreset Selection, Graph Coarsening, Graph Condensation, and Ego-Graph Coresets. Each paradigm offers distinct principles and design choices, reflecting different trade-offs between computational efficiency, structural preservation, and task-specific performance.

Classical Coreset Selection. Classical coreset selection techniques, widely studied in the context of independent and identically distributed (i.i.d.) data, provide a foundation for reducing training complexity, though they often fail to account for the relational dependencies inherent in graph domains. Among the simplest is *Random Sampling*, which selects data points uniformly at random. While computationally trivial and a useful baseline, it disregards both feature diversity and task relevance, often leading to unstable or biased subsets. A more structured alternative is *Herding* Welling (2009), which greedily selects points approximating the mean of each class distribution in feature space. This ensures coverage of prototypical examples but tends to neglect boundary cases and minority samples that are crucial for refining decision boundaries. Similarly, *K-Center* Farahani and Hekmatfar (2009); Sener and Savarese (2017) minimizes the maximum distance from any training point to the nearest selected center, thus enforcing broad coverage. However, its reliance on Euclidean distances assumes homogeneous feature geometry and ignores the relational structure that drives GNN learning. Beyond these commonly used strategies, several other classical methods have shaped the field. *Facility Location* and *K-Medoids* methods Mirzasoleiman et al. (2020) aim to identify exemplars that maximize representativeness while minimizing redundancy, often through submodular optimization frameworks. *Clustering-based approaches*, such as K-Means or spectral clustering, group data into partitions and select representative points from each cluster to balance diversity and density. More recently, *Gradient-based coreset selection* Sener and Savarese (2017); Killamsetty et al. (2021) has gained traction by leveraging the alignment of gradient information between the coreset and the full dataset, directly linking selection to model optimization. While these techniques achieve remarkable performance in standard classification tasks, they share a critical limitation in the graph setting: they assume independence between samples and therefore overlook the importance of local topological context and message-passing dynamics. Consequently, when applied directly to GNNs, their subsets may be feature-diverse yet structurally myopic, restricting their ability to support robust graph representation learning.

Graph Coarsening. Graph coarsening techniques aim to construct reduced graphs that preserve essential structural, spectral, or task-relevant properties of the original graph, enabling efficient training of graph neural networks. Unlike node-level sampling, coarsening hierarchically aggregates nodes into supernodes or regions, often guided by connectivity, label information, or signal-propagation dynamics. One prominent approach, *Universal Graph Coarsening (UGC)* Kataria et al. (2024), formulates coarsening as a self-supervised learning problem, optimizing a coarsening operator to align node embeddings between the original and reduced graphs. UGC's key strength lies in its adaptability to both homophilic and heterophilic graphs, enabling generalization across domains. However, its reliance on spectral similarity may struggle when global alignment conflicts with local task signals. Another paradigm is *Message-Passing Coarsening (MP-Coarsening)* Joly and Keriven (2024), which explicitly preserves the message-passing dynamics of GNNs during the coarsening process. By constructing coarsened graphs that ensure consistent neighborhood aggregation, MP-Coarsening offers theoretical guarantees for representation fidelity across layers. This principled design, however, comes at the cost of increased computational complexity, especially in dense graphs or high-resolution hierarchies, and may require fine-tuning to balance coarsening depth with training efficiency. In contrast, *Supervised Granular-Ball Graph Coarsening (SGBGC)* Xia et al.

(2025) incorporates label information directly into the coarsening procedure by iteratively merging structurally similar nodes that also share semantic proximity. This integration of structural and task-awareness improves discriminative power in the reduced graph but introduces a strong dependency on label quality and distribution. Moreover, its clustering process can be sensitive to hyperparameters and lacks theoretical guarantees for preserving spectral properties. Collectively, these coarsening methods offer scalable alternatives to full-graph GNN training, yet often involve trade-offs between structural fidelity, task relevance, and computational efficiency. While UGC emphasizes spectral consistency, MP-Coarsening enforces propagation alignment, and SGBGC injects supervised signals during clustering, each with distinct assumptions and limitations that motivate the development of more localized and task-aware selection strategies like TAS-EGNN.

Graph Condensation. Graph condensation techniques aim to generate compact synthetic graphs that faithfully preserve the training dynamics of the original large-scale graphs. A seminal approach in this domain is *GCond (Graph Condensation for GNNs)* Jin et al. (2021), which formulates condensation as a bi-level optimization problem: it synthesizes a small graph by matching the gradients of a GNN trained on this synthetic graph with those on the full dataset. This gradient-matching strategy enables the distilled graph to mimic the original's learning trajectory, enabling efficient training while retaining predictive performance. However, this method often requires substantial computational resources during condensation, particularly due to the backpropagation of meta-gradients and the practical instability of bilevel optimization. Building on the limitations of gradient-based condensation, *Class-Partitioned Graph Condensation (CGC)* Gao et al. (2025) proposes a training-free alternative that avoids gradient computation altogether. Instead, CGC partitions nodes into class-specific clusters based on structure and label priors, then selects representative nodes to construct a condensed graph. This shift drastically reduces condensation time and enhances scalability. Nevertheless, its reliance on heuristic clustering and class priors can lead to suboptimal generalization when the underlying class distribution or graph topology is noisy or heterogeneous. While graph condensation techniques offer a promising approach to dataset reduction, they often face trade-offs among computational cost, fidelity to the original training dynamics, and robustness to structural noise.

Ego-Graph Coresets. A recent line of work selects *ego-graphs*, localized neighborhoods, as the basic training units. *Spectral Greedy Graph Coresets (SGGC)* Ding et al. (2024) is a two-stage method: (i) a coarse "spectral coverage" stage uses GIGA to pick widely separated centers by maximizing an upper bound independent of per-ego spectral embeddings, thereby approximating the full-graph spectral embedding without evaluating it on each ego-graph; (ii) a refinement stage applies submodular maximization to filter candidates whose *spatial* embeddings are deemed non-distinct, improving sample efficiency. SGGC also compresses diffusion ego-graphs via low-rank (PCA) feature compression to control data size. This design promotes structural diversity while remaining model-agnostic and scalable. However, the pipeline rests on a "small variation" assumption for spectral embeddings on ego-graphs. It reports an overall time complexity of $O(cn_t n)$, which can be constraining at larger budgets and scales. These choices motivate alternatives that retain structural coverage yet align selection more directly with supervised task signals.

Limitations of Prior Work. Despite significant progress, prior approaches each suffer from inherent limitations. Classical coreset methods, while computationally lightweight, overlook graph structure and thus fail to capture relational dependencies crucial for GNNs. Graph coarsening strategies provide efficient reductions and preserve global structure but often ignore predictive uncertainty and supervised task signals, limiting their adaptability to dynamic learning scenarios. Condensation methods address task alignment more directly but typically rely on expensive bi-level optimization or heuristic clustering, thereby limiting their scalability to large graphs. Within ego-graph coresets, SGGC narrows the task–structure gap but still relies on spectral coverage plus a submodular *topology distinctiveness* filter rather than explicit supervision, and its "small variation" assumption may weaken on very dense graphs or non-message-passing architectures. In addition, the time complexity is a practical limitation of the SGGC method.

In contrast, our proposed TAS-EGNN advances this landscape by introducing a principled, task-aware scoring framework that unifies three complementary signals: spectral entropy to capture local structural heterogeneity, predictive entropy to quantify model uncertainty, and misclassification feedback to highlight challenging training examples. These signals are integrated through a greedy coverage algorithm designed to maximize informativeness while reducing redundancy across ego-graphs. By coupling localized structural analysis with supervision-aware signals, TAS-EGNN achieves a balance between scalability and precision, enabling robust coreset selection for

both medium-scale and large-scale graph learning without the computational overhead of eigendecomposition or condensation pipelines.

B Settings and Additional Experiments

Datasets. We evaluate TAS-EGNN on three graph families. (1) *Graph Benchmarks*: **Cora** and **Citeseer** Kipf and Welling (2016) are two classical citation networks with bag-of-words features, **Flickr** Zeng et al. (2020) is a social network dataset encoding user attributes and relations, and **OGBN-Arxiv** and **OGBN-Products** Hu et al. (2020) are a large-scale graph dataset from the Open Graph Benchmark. (2) *Transaction-fraud detection benchmarks*: **Banksim** Lopez-Rojas and Axelsson (2014), **Paysim** Lopez-Rojas et al. (2016), and **ECC** Dal Pozzolo et al. (2014), where nodes are accounts (or counterparties), edges connect accounts via observed transactions, node features aggregate transactional statistics, and labels are binary (normal vs. fraud). Table 6 summarizes the characteristics of all the used datasets.

Table 6: Dataset statistics used in our experiments. Counts for fraud graphs reflect our transaction-to-graph construction (undirected).

Dataset	#Nodes	#Edges	#Features	#Classes	Positive-Rate
Cora	2,708	5,429	1,433	7	/
Citeseer	3,327	4,732	3,703	6	/
Flickr	89,250	899,756	500	7	/
OGBN-Arxiv	169,343	1,166,243	128	40	/
OGBN-Products	2,449,029	123,718,152	100	47	/
Banksim	17,200	98,879	7	2	1.2%
Paysim	18,213	100,654	7	2	1%
ECC	20,000	108,800	30	2	2%

Implementation Details. All experiments were conducted using Python and PyTorch Geometric, running on an Intel Core $i9 - 14900 \times 32$ with an NVIDIA GeForce RTX 4070 GPU. Dataset splits followed the standard settings provided in prior work, ensuring consistency with baseline comparisons. To evaluate the adaptability of TAS-EGNN across different message-passing paradigms, we integrated it with five representative GNN backbones: Graph Convolutional Network (GCN) Kipf and Welling (2016), Graph Attention Network (GAT) Veličković et al. (2018), Graph Isomorphism Network (GIN) Xu et al. (2018), Approximate Personalized Propagation of Neural Predictions (APPNP) Gasteiger et al. (2018), and GraphSAGE Hamilton et al. (2017). For fairness, we adopted the same hidden dimension of 64 across all models. Training was performed using the Adam optimizer with a learning rate of 0.01 and weight decay of 5×10^{-4}. Epochs were set to 200 for Flickr, OGBN-Arxiv, and OGBN-Products, and 300 for Cora, Citeseer, Banksim, Paysim, and ECC, with early-stopping patience fixed at 10. Accuracy was used as the primary evaluation metric for the Citation, Social, and products datasets, and PR-AUC for the Transaction-fraud detection datasets; all reported results represent the mean and standard deviation across 10 independent runs.

Warm-up Model for Computing Task Signals. To compute the predictive uncertainty $\mathcal{H}_{\text{pred}}(v)$ and the misclassification indicator $\mathbb{1}[\hat{y}_v \neq y_v]$, we first train a standard backbone model f_θ on the full training split. This warm-up model is optimized using cross-entropy loss and early stopping based on validation accuracy, and is used solely to produce logits and softmax probabilities for computing the two task-aware signals. After coreset selection, a fresh backbone model is reinitialized and trained from scratch exclusively on the induced subgraph. For the fraud datasets in the one-class setting, this warm-up model is trained only on normal-class nodes. We found that this warm-up initialization yields stable uncertainty estimates and reliably identifies hard-to-classify samples.

Baselines. We consider three classical coreset selection techniques: **Random Sampling**, **Herding** Welling (2009), and **K-Center** Farahani and Hekmatfar (2009); Sener and Savarese (2017), all of which operate on i.i.d. assumptions and disregard graph topology. We also evaluate two recent **graph condensation** methods: **GCond** Jin et al. (2021), which employs gradient matching to synthesize condensed graphs, and **CGC** Gao et al. (2025), a training-free approach that clusters nodes by class prototypes. In the **graph coarsening** category, we include **SGBGC** Xia et al. (2025), which forms granular balls based on structural and label proximity, and **UGC** Kataria

et al. (2024), a universal coarsening method that preserves spectral properties across graph types. Finally, for **ego-graph coresets**, we use **SGGC** Ding et al. (2024), which greedily selects subgraphs using spectral diversity and classifier alignment. All baseline results follow the authors' official implementations or faithfully reproduce their reported settings.

B.1 Backbone GNN Selection

Settings. Before comparing TAS-EGNN against the baseline techniques, we first evaluate its performance under different backbone GNN architectures. The goal of this evaluation is to identify the most suitable GNN for subsequent experiments, ensuring a fair and representative assessment of our framework. We integrate TAS-EGNN with five representative GNN models: GCN Kipf and Welling (2016), GAT Veličković et al. (2018), GIN Xu et al. (2018), APPNP Gasteiger et al. (2018), and GraphSAGE Hamilton et al. (2017). These architectures span a spectrum of message-passing paradigms, including spectral convolutions, attention-based aggregation, and neighborhood sampling, allowing us to test TAS-EGNN's adaptability across diverse modeling strategies. Training protocols, optimizer configurations, and early stopping were kept consistent across all backbones to isolate the effect of the GNN choice.

Results and Discussion. The classification accuracy results are reported in Table 7. Several important observations can be made. First, across all four datasets, GraphSAGE consistently delivers the strongest or near-strongest results when coupled with TAS-EGNN. For instance, at a 50% coreset ratio on Cora, GraphSAGE achieves 90.4% accuracy, surpassing GCN (86.8%), GAT (86.6%), and APPNP (85.9%). Similarly, in large-scale datasets such as Flickr, OGBN-Arxiv, and OGBN-Products, GraphSAGE demonstrates superior scalability and robustness, achieving 75.8%, 68.6%, and 96.5 accuracy at 2.0%, 2.0%, and 0.2% coreset ratios, respectively. On transaction-fraud datasets (PR-AUC), GraphSAGE again dominates: Banksim achieves 96.3%, 96.4%, and 96.5%; Paysim achieves 94.6%, 96.0%, and 96.1%; and ECC achieves 97.4%, 97.5%, and 97.6% at budgets of 15%, 25%, and 50%, respectively. Second, while GCN and GAT perform competitively on smaller citation networks, their gains diminish on large-scale datasets, where local neighborhood sampling and inductive generalization become more critical. APPNP, though theoretically strong due to personalized PageRank propagation, underperforms relative to GraphSAGE, particularly in sparse or heterogeneous settings such as Flickr. GIN achieves moderate accuracy but suffers from instability at low coreset ratios, likely due to its sensitivity to local degree distributions. The Δ column, defined as the margin between the best and second-best backbone under the same TAS-EGNN selection, quantifies how decisively the top model wins. On Flickr, Δ is consistently large (about 15-16 points across budgets), indicating that TAS-EGNN coresets pair exceptionally well with GraphSAGE in sparse, heterogeneous settings. Cora shows robust margins ($\Delta \approx 5$ at all budgets), and OGBN-Arxiv and OGBN-Products exhibit steady advantages ($\Delta \approx$ 3-4), reinforcing the scalability of the TAS-EGNN+GraphSAGE combination on larger graphs. Citeseer is mixed: a strong lead at 15% ($\Delta = 6.8$), a near-tie at 25% where APPNP slightly leads ($\Delta = 0.2$), and a clear win again at 50% ($\Delta = 3.2$). For the fraud datasets, Δ is modest on *Banksim* (1.3-2.3) and small on *Paysim* (0.5-0.6), suggesting a ceiling effect in PR-AUC under balanced coresets, while *ECC* shows larger gaps (3.8-5.6), pointing to meaningful benefits from TAS-EGNN's task-aware selection. Collectively, these gaps show that TAS-EGNN not only lifts overall accuracy and PR-AUC but does so with *decisive* gains on harder, larger, or more heterogeneous graphs, and at least parity in saturated regimes. The consistent outperformance of GraphSAGE suggests that its inductive nature and effective neighborhood aggregation make it especially well-suited to complement TAS-EGNN's ego-graph-based coreset strategy. By explicitly combining local structural diversity with task-aware scoring, TAS-EGNN leverages GraphSAGE's ability to generalize across unseen neighborhoods, particularly in large-scale graphs. Based on these findings, we adopt GraphSAGE as the default backbone for all subsequent comparisons against baseline coreset and condensation methods.

B.2 Hyperparameter Sensitivity

Sensitivity to Scoring Weights (α, β, γ). To assess the robustness of TAS-EGNN to the choice of scoring coefficients (α, β, γ), we conducted a sensitivity experiment on two representative datasets: Cora and Banksim. For Cora, performance was evaluated using node classification accuracy, while for Banksim, we used PR-AUC consistent with the fraud-detection evaluation protocol. In each run, we varied one coefficient within $\{0.0, 0.25, 0.5, 0.75, 1.0\}$ while fixing the other two at 1.0, thereby isolating the influence of each score component. We used GraphSAGE as the backbone encoder, a coreset budget of 50%, and averaged results over 10 runs. The table accompanying this analysis reflects these controlled variations and enables direct comparison of performance trends with respect

Table 7: GNNs Classification performance comparison using TAS-EGNN in terms of accuracy metric on Citation, Social, and Products datasets, and PR-AUC metric on Transaction-Fraud Detection datasets. Best in bold. Δ is the gap between the first and the second best GNN's performance.

Dataset	Ratio	APPNP	GAT	GCN	GIN	GraphSAGE	Δ
Citeseer	15%	56.6±0.4	53.6±0.6	57.2±0.5	58.6±1.6	**65.4±1.2**	6.8
	25%	**71.9±1.6**	67.1±1.5	71.5±0.3	60.2±1.3	71.7±0.9	0.2
	50%	75.7±0.3	72.7±1.5	77.1±0.4	63.6±1.9	**80.3±0.5**	3.2
Cora	15%	73.3±0.9	62.4±0.5	69.4±0.2	78.8±1.5	**83.9±0.1**	5.1
	25%	82.4±0.1	81.6±0.3	82.8±0.1	80.3±1.2	**87.9±0.1**	5.1
	50%	85.9±0.4	86.6±0.3	86.8±0.1	87.0±0.8	**92.3±0.2**	5.3
Fliker	0.5%	51.3±0.9	52.4±2.3	56.8±0.3	53.7±1.6	**72.9±1.2**	16.1
	1.0%	52.4±0.5	54.7±0.8	58.9±0.3	55.5±2.1	**74.4±1.7**	15.5
	2.0%	52.4±0.3	54.8±1.1	59.7±0.2	55.7±1.0	**75.8±1.6**	16.1
OGBN-Arxiv	0.5%	51.1±0.2	59.4±0.4	62.5±0.2	59.0±0.3	**66.6±0.2**	4.1
	1.0%	62.9±0.3	60.9±0.4	64.3±0.1	59.9±1.0	**67.4±0.2**	3.1
	2.0%	63.3±0.3	61.6±0.7	65.1±0.2	60.8±1.0	**68.6±0.1**	3.5
OGBN-Products	0.1%	59.4±0.5	60.9±0.4	61.1±0.3	58.8±0.3	**64.8±0.5**	3.7
	0.2%	66.2±0.6	64.4±0.5	65.2±0.4	62.7±0.3	**69.5±0.6**	3.3
Banksim	15%	95.0±0.3	93.7±1.3	92.8±3.3	91.3±0.4	**96.3±0.2**	1.3
	25%	94.9±0.4	93.9±0.8	93.0±2.5	91.7±0.4	**96.4±0.2**	1.5
	50%	94.2±0.6	94.1±0.4	93.1±1.5	92.2±0.3	**96.5±0.5**	2.3
Paysim	15%	92.7±1.1	88.7±0.5	94.0±1.6	89.1±0.6	**94.6±1.1**	0.6
	25%	93.3±2.5	88.9±0.8	95.4±2.6	89.4±0.5	**96.0±1.0**	0.6
	50%	93.7±0.5	89.2±0.6	95.6±0.9	89.9±0.3	**96.1±1.4**	0.5
ECC	15%	91.1±0.9	87.5±0.4	91.8±0.7	88.0±0.3	**97.4±0.5**	5.6
	25%	92.2±0.2	87.9±0.3	92.4±0.9	88.4±0.2	**97.5±0.6**	5.1
	50%	93.4±0.1	88.3±0.1	93.8±0.6	88.6±0.5	**97.6±0.2**	3.8

to the weighting of spectral entropy, predictive uncertainty, and supervised misclassification feedback.

The results show smooth and monotonic performance progression as weights increase, with no sharp fluctuations or instability. Increasing α consistently boosts performance on both datasets, especially on Banksim, supporting the importance of structural heterogeneity in irregular graph topologies. Increasing β yields steadily improving accuracy on Cora, indicating that uncertainty-driven selection is particularly informative in citation graphs with well-separated class clusters. Varying γ produces modest and stable gains, confirming that misclassification serves as a supplementary supervisory cue rather than a dominant driver. Importantly, across all sweeps, performance remains within a narrow variance band, demonstrating that TAS-EGNN is robust under broad weighting changes and does not require fine-grained tuning. Based on these observations, we recommend the balanced default setting $(\alpha, \beta, \gamma) = (1.0, 1.0, 1.0)$, which consistently yielded strong and reliable performance across datasets.

Sensitivity to Diversity Coefficient (τ)**.** We additionally evaluate the influence of the diversity coefficient τ, which controls the penalization of ego-graph overlap during greedy selection. In this experiment, we vary $\tau \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$ while holding the other scoring components fixed at $(\alpha, \beta, \gamma) = (1.0, 1.0, 1.0)$. As before, accuracy is reported for Cora and PR-AUC for Banksim, averaged over 10 independent runs. This setup isolates the role of τ while ensuring consistent training and evaluation conditions.

The results in Table 9 show that moderate overlap penalization improves performance: $\tau = 0.25$ yields the best accuracy on Cora (92.3%) and the best PR-AUC on Banksim (96.5%). When $\tau = 0$, the model tends to repeatedly select ego-graphs from the same structural regions, resulting in reduced coverage diversity. Conversely, large values of τ (e.g., 0.75 or 1.0) over-emphasize diversity and may exclude task-relevant but spatially clustered

Table 8: Sensitivity of TAS-EGNN to the score weighting parameters (α, β, γ), with accuracy as the evaluation metric for Cora and PR-AUC for Banksim. Only one parameter varies per block, while the remaining two are fixed to 1.0. Results are reported as mean ± standard deviation over 10 runs.

			$\beta, \gamma = 1.0$		
α	0.0	0.25	0.5	0.75	1.0
Cora	91.9±0.3	92.0±0.1	92.1±0.2	92.2±0.1	92.3±0.1
Banksim	93.7±1.1	94.1±0.6	94.7±0.8	95.8±0.7	96.5±0.5
			$\alpha, \gamma = 1.0$		
β	0.0	0.25	0.5	0.75	1.0
Cora	90.4±0.1	90.8±0.2	91.3±0.1	91.9±0.2	92.3±0.1
Banksim	95.6±0.4	95.8±0.3	96.1±0.4	96.3±0.3	96.5±0.5
			$\alpha, \beta = 1.0$		
γ	0.0	0.25	0.5	0.75	1.0
Cora	92.2±0.1	92.2±0.2	92.2±0.2	92.3±0.1	92.3±0.1
Banksim	96.0±0.3	96.2±0.4	96.3±0.5	96.4±0.6	96.5±0.5

Table 9: Sensitivity of TAS-EGNN to the diversity coefficient τ, reporting accuracy on Cora and PR-AUC on Banksim. Results are mean ± std across 10 runs.

τ	0.0	0.25	0.5	0.75	1.0
Cora	92.0±0.2	92.3±0.1	92.1±0.1	91.7±0.2	91.2±0.1
Banksim	95.7±0.6	96.5±0.5	96.1±0.4	95.2±0.5	94.8±0.3

regions, leading to a slight decline in performance. These results confirm that τ plays a key role in balancing informativeness versus redundancy, and empirically support the use of $\tau = 0.25$ as a robust default that yields consistent improvements across datasets.