WAVEGUARD: COORDINATING MESSAGE PASSING AND TEST-TIME ADAPTATION WITH SPECTRAL MEMORY

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph test-time adaptation (GTTA) updates a pretrained graph neural network using an unlabeled target graph without revisiting the source graph. Under structural shift, altered neighborhoods can both degrade node representations and corrupt the signals used for adaptation. Predictive confidence alone does not distinguish these effects, particularly when a model is confidently wrong. We propose WaveGuard, which uses a fixed node-wise intervention score to coordinate message passing and unsupervised updates. A compact class–degree-conditioned memory of source graph-wavelet energy provides a reference for target spectral deviation. Combining this deviation with frozen-model uncertainty yields a score that controls direction-aware soft topology purification and calibrates entropy, anchor, and consistency losses. The score, reweighted graph, and prediction anchors remain fixed during adaptation, while deployment retains only aggregate source statistics. Experiments on five graph benchmarks covering 23 target scenarios demonstrate WaveGuard's effectiveness under structural, temporal, and mixed shifts. WaveGuard ranks first in six of eight CSBM scenarios, achieves the highest average accuracy on ogbn-arxiv, and outperforms the evaluated baselines in all three PubMed settings and two of three Twitch transfer tasks. On Elliptic, it surpasses the strongest baseline by 6.12 and 6.63 percentage points in Balanced Accuracy and Macro-F1, respectively. Edge-weight analysis and component ablations across seven scenarios support the complementary contributions of message control and objective calibration.

1 INTRODUCTION

Graph neural networks (GNNs) support relational tasks in academic networks, social platforms, and financial fraud detection, but their reliance on neighborhood aggregation makes deployment sensitive to changes in graph structure (Mao et al., 2023; Platonov et al., 2023; Luan et al., 2023; Li et al., 2025). Real graphs evolve across time, regions, and platforms. Changes in degrees, homophily, and local connectivity alter the information reaching each node, even when its own attributes remain unchanged. Graph domain adaptation and out-of-distribution (OOD) generalization address such shifts, but typically require simultaneous source access, multiple training environments, or retraining (Li et al., 2025; Liu et al., 2024a). Graph test-time adaptation (GTTA) instead adapts a pretrained model to an unlabeled target graph when source-node records are no longer available (Liang et al., 2025; Jin et al., 2023; Mao et al., 2024).

Structural shifts challenge GTTA through two coupled effects. Semantically inconsistent neighborhoods reduce the separability of hidden representations, and those representations then generate the predictions used by unsupervised adaptation. Entropy minimization can reinforce confident errors, while consistency between perturbed views can preserve the same neighborhood-induced mistake. Existing methods address parts of this problem through graph transformation, neighborhood augmentation, aggregation adaptation, or target-side objectives (Jin et al., 2023; Ju et al., 2023; Mao et al., 2024; Bao et al., 2025; Hsu et al., 2026). A node-wise reference for deciding how strongly to intervene could connect these operations: nodes affected by a structural shift should contribute differently to both incoming messages and parameter updates.

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.

Multiscale spectral responses provide such a reference. In Figure 1, heterophilic rewiring and OOD-hub rewiring increase high-frequency mass and spectral centroid, while a homophily-preserving control does not. Their band-energy profiles also differ at 20% rewiring. Graph wavelets measure these changes locally and across scales, allowing target nodes to be compared with source structural statistics (Defferrard et al., 2016; Hammond et al., 2011). Conditioning this comparison on class and degree is essential because normal hubs and class-boundary nodes can already exhibit distinct spectral responses without a distribution shift.

We propose WaveGuard to translate this source reference into a fixed intervention score. Offline, WaveGuard summarizes source hidden representations as class–degree-conditioned graph-wavelet statistics. At deployment, it combines target spectral deviation and frozen-model uncertainty, then uses the resulting score to attenuate directional messages and calibrate an unsupervised objective. Larger scores induce stronger message control and favor reliable frozen prediction anchors; smaller scores favor entropy minimization and consistency learning. Fixing the score and anchors prevents changing confidence during adaptation from redefining the control signal. Only compact source summaries are retained, with no source-node replay or target labels.

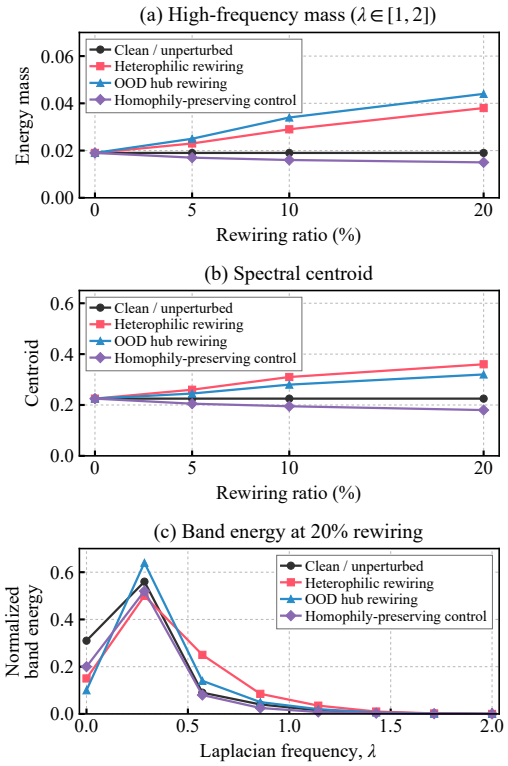

Figure 1: Spectral responses to structural perturbations. Harmful rewiring increases high-frequency energy, unlike the homophily-preserving control.

Our contributions are:

- A source-referenced intervention score that combines class–degree-conditioned multiscale spectral deviation with frozen-model uncertainty for structure-dominated GTTA.
- A dual intervention mechanism in which the same fixed score controls directional message weights and node contributions to entropy, anchor, and consistency losses.
- Evaluation on five benchmarks and 23 target scenarios, together with edge selectivity, adaptation dynamics, and ablations that characterize when the two interventions help and where mixed shifts remain challenging.

2 Related Work

Test-time adaptation on graphs. Generic TTA adapts pretrained models through entropy minimization, pseudo-labels, or consistency objectives (Liang et al., 2025; Wang et al., 2021; Iwasawa & Matsuo, 2021; Boudiaf et al., 2022). Graph-specific methods additionally exploit or modify connectivity. GTrans refines target features and structure (Jin et al., 2023); GraphPatcher generates virtual nodes for deficient neighborhoods (Ju et al., 2023); and SOGA combines information maximization with structural consistency (Mao et al., 2024). Matcha adapts hop-aggregation parameters using prediction-informed clustering (Bao et al., 2025), while TSA combines uncertainty-aware neighborhood weighting with adaptive self-node and neighborhood aggregation (Hsu et al., 2026). WaveGuard uses compact source spectral statistics to construct one fixed control signal for both directional messages and unsupervised updates. The comparison therefore concerns the source reference and coupling of interventions, as well as which target-side quantities each method changes.

Domain adaptation and spectral learning. Graph domain adaptation typically estimates discrepancies using both labeled source and unlabeled target graphs (Liu et al., 2023; 2024a;b; You et al., 2023; Cai et al., 2024). Graph OOD generalization instead improves unseen-domain behavior through source-side augmentation, extrapolation, or representation learning (Sui et al., 2023; Li et al., 2024a;b). WaveGuard operates with a different information budget: a pretrained model and aggregate source summaries. Graph filters and wavelets encode smoothness and localized frequency variation (Xu et al., 2024a;b; Defferrard et al., 2016; Hammond et al., 2011); here they define a deployment-time reference for structural deviation. Appendix E gives the extended comparison across information access, spectral conditioning, and intervention mechanisms.

3 Problem Setup and Structural Motivation

Adaptation setting. We consider transductive source-free GTTA for node classification. A GNN with parameters θ_0 is pretrained on a labeled source graph $\mathcal{G}^s$ and adapts to an unlabeled target graph $\mathcal{G}^t = (\mathbf{X}^t, \mathbf{A}^t)$. Deployment provides the target graph, the pretrained model, and compact source statistics; target labels and source-node features, labels, and representations remain unavailable. WaveGuard reweights the existing edge support and updates a prespecified subset of parameters. Its memory scales with the numbers of classes, degree buckets, and spectral scales, rather than the number of source nodes. We target degree, homophily, local-rewiring, and temporal structural shifts; attribute and label shifts may coexist. Appendix A.1–A.2 formalizes the shift definitions and information budget.

Why coordinate messages and updates? Structural shifts change which hidden representations are compared and aggregated, redistributing energy across graph frequencies. Multiscale graph wavelets capture this redistribution from local perturbations to broader propagation changes, and a sparse Chebyshev approximation avoids eigendecomposition. These measurements complement predictive entropy: confidence can remain high after a neighborhood-induced error, and two perturbed views can agree on the same error. Reweighting the loss alone leaves harmful messages active, whereas changing topology alone does not determine which predictions provide reliable adaptation signals. This motivates a common node-wise intervention score for both operations. Appendix A.3–A.4 provides the spectral definitions and analysis of standard unlabeled objectives.

4 WaveGuard

Figure 2 summarizes the method. A frozen encoder first produces hidden representations and predictions on the original target graph. Source-referenced spectral deviation and predictive uncertainty define a fixed intervention score, which determines a softly reweighted graph and node-specific objective weights. Frozen predictions on the original and reweighted graphs supply anchors for subsequent adaptation. The evaluated GSN, GCN, and GPRGNN backbones expose the hidden representations and message weights required by these operations (Kipf & Welling, 2017; Hamilton et al., 2017; Chien et al., 2021).

4.1 A Source-Referenced Intervention Score

Multiscale representations. Predictive entropy compresses the effect of a shifted neighborhood into class probabilities. WaveGuard also measures the frozen pre-classifier representation, retaining structural information before that compression. For node features $\mathbf{X}$ and adjacency $\mathbf{A}$, the encoder produces

$$\mathbf{H} = f_{\theta_0}^{\text{enc}}(\mathbf{X}, \mathbf{A}) = [\mathbf{h}_1, \ldots, \mathbf{h}_N]^\top \in \mathbb{R}^{N \times d_h}. \tag{1}$$

Given the symmetric normalized Laplacian $\mathbf{L}$, the high-pass wavelet kernel and its order-R Chebyshev approximation yield

$$g_{s_k}(\lambda) = \lambda \exp(-s_k \lambda), \qquad \mathbf{Z}_k = g_{s_k}(\mathbf{L})\mathbf{H} \approx \sum_{r=0}^{R} a_{k,r} T_r(\mathbf{L})\mathbf{H}. \tag{2}$$

We use scales $\mathcal{S} = \{0.05, 0.1, 0.2, 0.5, 1.0\}$ and $R = 5$. Each response contributes a local energy measurement; logarithmic normalization limits sensitivity to differences in signal magnitude. The

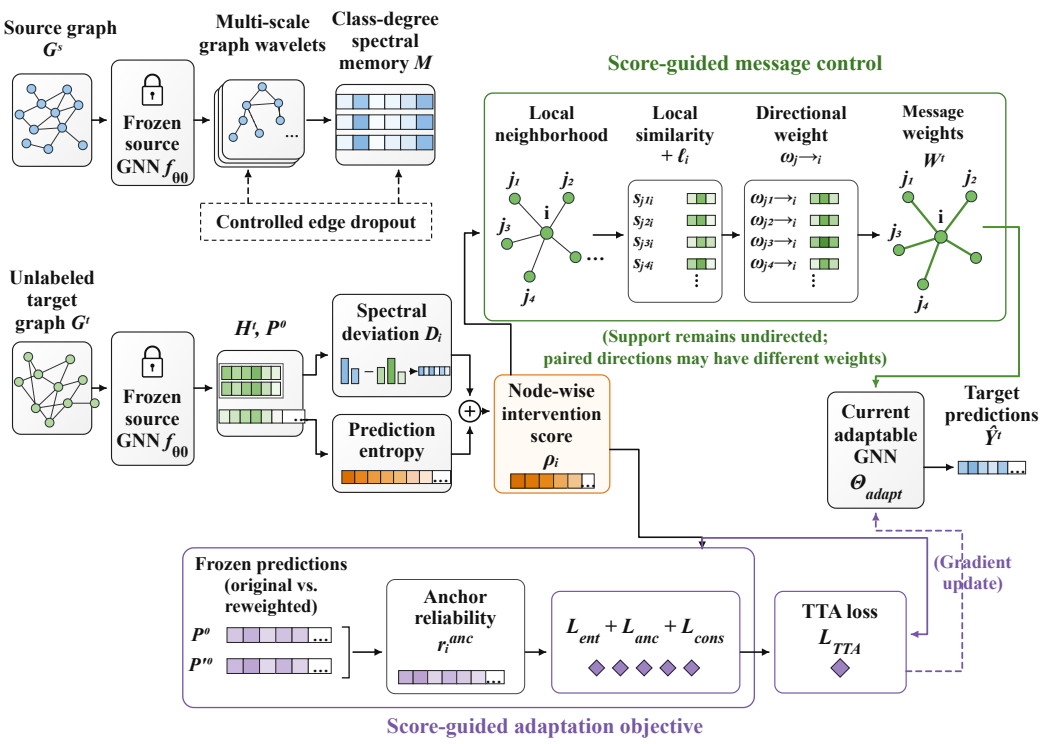

Figure 2: WaveGuard overview. Source spectral memory and target signals define a fixed intervention score for message control (green) and objective calibration (purple). Figure F1 enlarges both branches; Appendix F explains the diagram notation.

resulting vector is the node's multiscale spectral signature:

$$\begin{aligned} E_i^{(k)} &= \| \mathbf{Z}_k[i,:] \|_2^2, \\ e_i^{(k)} &= \log\left(E_i^{(k)} + \epsilon\right), \\ \mathbf{e}_i &= \left[e_i^{(1)}, \ldots, e_i^{(K)}\right]^\top. \end{aligned} \tag{3}$$

Conditional source memory. Offline, source nodes are softly assigned to equal-frequency degree buckets. For each class–degree–scale condition, the memory stores the energy mean $\mu_{c,b,k}$, standard deviation $\sigma_{c,b,k}$, scale weight $w_{c,b,k}$, and a 95th-percentile normal-deviation threshold. Scale weights combine structural sensitivity under controlled source-edge dropout with class discriminability. The source GNN remains frozen, and unperturbed source nodes provide the retained reference statistics and thresholds. These aggregates distinguish normal class- and degree-dependent variation from target deviation without retaining source-node records. Appendix B.1 specifies the complete construction and fallback rules.

Target deviation and score. At deployment, each target energy vector is compared with the conditional source reference. Soft degree assignments $\alpha_{i,b}^t$ and frozen class probabilities $p_{i,c}^0$ combine the conditional deviations without committing to hard pseudo-labels:

$$D_i(c,b) = \sum_{k=1}^{K} w_{c,b,k} \left(\frac{e_{i,t}^{(k)} - \mu_{c,b,k}}{\sigma_{c,b,k} + \epsilon} \right)^2. \tag{4}$$

$$D_i(c) = \sum_{b=1}^{B} \alpha_{i,b}^t D_i(c,b), \qquad D_i = \sum_{c=1}^{C} p_{i,c}^0 D_i(c). \tag{5}$$

The corresponding source thresholds are marginalized in the same way to obtain q_i. Combining deviation with normalized frozen-model entropy $\mathcal{H}(\mathbf{p}_i^0)$ gives

$$\rho_i = \text{sigmoid}\left(\frac{D_i + \beta\mathcal{H}(\mathbf{p}_i^0) - q_i}{\tau_r}\right), \qquad 0 < \rho_i < 1. \tag{6}$$

where β weights predictive uncertainty and τ_r controls gating smoothness. The score specifies intervention strength, rather than a calibrated probability of corruption or prediction error. Class and degree conditioning reduce the influence of normal structural heterogeneity; uncertainty complements the spectral reference when class predictions are ambiguous. We compute the score once on the original target graph and hold it fixed throughout adaptation. Appendix B.2 gives the entropy and threshold definitions.

4.2 Directional Message Control

For each aggregation direction $j \to i$ of an undirected supporting edge, WaveGuard compares frozen hidden representations with a receiver-specific similarity quantile. The resulting inconsistency gate $m_{j\to i}$ is larger when a message is unusually dissimilar within the receiving node's neighborhood. Dissimilarity alone is insufficient because legitimate heterophilic edges can connect different representations. We therefore combine the gate with sender and receiver scores, then convert the resulting intervention into an edge weight:

$$\begin{aligned} r_{j\to i} &= 1 - (1-\rho_j)\,(1-\rho_i m_{j\to i})\,, \\ \omega_{j\to i} &= 1 - (1-\omega_{\min}) r_{j\to i}, \\ W^t_{ji} &= A^t_{\text{ji}}\omega_{j\to i}. \end{aligned} \tag{7}$$

A high sender score attenuates outgoing messages; with a lower sender score, attenuation depends on both the receiver score and local inconsistency. The lower bound $\omega_{\min} > 0$ preserves the original support, and self-loops are retained. The original adjacency $\mathbf{A}^t$ remains symmetric; the directed message-weight matrix $\mathbf{W}^t$ may be asymmetric. We normalize these weights for propagation and keep them fixed during adaptation. Appendix B.3 defines the local quantile and directional gate.

4.3 Score-Calibrated Adaptation

Reliable prediction anchors. On the reweighted graph, the current prediction $\mathbf{p}_i^\theta$ is adapted using frozen predictions from the original and reweighted graphs. Their Jensen–Shannon discrepancy measures sensitivity to purification, while the reweighted prediction entropy measures uncertainty. Together they determine whether the frozen reweighted prediction supplies a reliable anchor:

$$\begin{aligned} \delta_i &= \text{JSD}\left(\mathbf{p}_i^{\text{orig}}, \mathbf{p}_i^{\text{pur}}\right), \\ r_i^{\text{anc}} &= \exp\left(-\frac{\delta_i}{\tau_a}\right)\left[1 - \mathcal{H}\left(\mathbf{p}_i^{\text{pur}}\right)\right]. \end{aligned} \tag{8}$$

Lower-score nodes receive more entropy and consistency weight. Higher-score nodes receive more anchor weight only when the anchor is reliable, with exponent γ controlling the separation:

$$w_i^{\text{ent}} = (1-\rho_i)^\gamma, \qquad w_i^{\text{anc}} = \rho_i^\gamma r_i^{\text{anc}}. \tag{9}$$

Unsupervised objective. The entropy term sharpens predictions primarily at lower-score nodes. The anchor term constrains drift at higher-score nodes through a reliable frozen target. Their losses are

$$\begin{aligned} \mathcal{L}_{\text{ent}} &= \frac{1}{N_t}\sum_{i=1}^{N_t} w_i^{\text{ent}}\mathcal{H}(\mathbf{p}_i^\theta), \\ \mathcal{L}_{\text{anc}} &= \frac{1}{N_t}\sum_{i=1}^{N_t} w_i^{\text{anc}}\,\text{KL}\left(\mathbf{p}_i^{\text{pur}} \parallel \mathbf{p}_i^\theta\right). \end{aligned} \tag{10}$$

Two independent dropout views provide an additional consistency signal. We use the view with the larger maximum class probability as the stop-gradient teacher, denoted $\mathbf{p}_i^{\text{hi}}$, and the other as $\mathbf{p}_i^{\text{lo}}$.

Applying the same lower-score weighting gives the consistency term and the complete objective:

$$\mathcal{L}_{\text{cons}} = \frac{1}{N_t} \sum_{i=1}^{N_t} w_i^{\text{ent}} \, \mathrm{KL}\left(\mathrm{sg}[\mathbf{p}_i^{\text{hi}}] \parallel \mathbf{p}_i^{\text{lo}}\right). \tag{11}$$

$$\mathcal{L}_{\text{TTA}} = \lambda_{\text{ent}}\mathcal{L}_{\text{ent}} + \lambda_{\text{anc}}\mathcal{L}_{\text{anc}} + \lambda_{\text{cons}}\mathcal{L}_{\text{cons}}. \tag{12}$$

Only the selected model parameters are updated. Score estimation, graph reweighting, and anchor construction occur once, so increasing confidence cannot itself relax a node's assigned intervention. The source memory occupies $O(CBK + B)$ space, independent of source-graph size. Appendix B.4–B.5 provides the full objective and complexity, Algorithm 1 gives the procedure, and Appendix C analyzes an idealized fixed-score surrogate objective.

5 Experiments

5.1 Experimental Setup

We evaluate WaveGuard on CSBM, ogbn-arxiv, Elliptic, PubMed, and Twitch, covering 23 target scenarios with controlled mixed shifts, temporal evolution, structural contamination, and cross-lingual transfer (Deshpande et al., 2018; Hu et al., 2020; Weber et al., 2019; Yang et al., 2016; Rozemberczki et al., 2021). We report accuracy except on class-imbalanced Elliptic, where both Balanced Accuracy and Macro-F1 are used. Each target starts independently from the same source checkpoint, and target labels are used only for evaluation. Main comparison results are means and standard deviations over five seeds. Backbones are GSN for CSBM, GCN for ogbn-arxiv, and GPRGNN for the remaining datasets; Appendix D.1–D.2 records partitions, parameter update scopes, training configurations, and hyperparameters.

The comparisons include generic TTA methods TENT, T3A, LAME, and ActMAD, together with graph methods GTrans, GraphPatcher, SOGA, Matcha, TSA, HomoTTT, GOAT, T3R, and TTReFT (Jin et al., 2023; Ju et al., 2023; Mao et al., 2024; Bao et al., 2025; Zhang et al., 2025; Hsu et al., 2026; Mirza et al., 2023; Wang et al., 2021; Iwasawa & Matsuo, 2021; Boudiaf et al., 2022; Zhang et al., 2024a; Truong et al., 2026; Zhang et al., 2026). Methods share dataset preprocessing, source checkpoints, target masks, and evaluation metrics; any method-specific source-side training follows the original method. Hyperparameters are selected using source validation or an independent development domain, without target evaluation labels. We first examine classification performance, then use edge selectivity, adaptation trajectories, and component removal to assess the two interventions.

5.2 Performance under Structural and Mixed Shifts

Synthetic shifts. Figure 3 compares representative methods on eight CSBM scenarios. WaveGuard reaches 91.2% ± 1.93% on CSBM1 and 88.6% ± 2.15% on CSBM3, with approximately 79.5% mean accuracy across all eight scenarios. Among the methods shown, it achieves the highest mean accuracy on CSBM1–3, CSBM5, CSBM7, and CSBM8, and the second-highest on CSBM4 and CSBM6. Gains over ERM range from 8.6 to 33.5 percentage points, with the largest improvements on CSBM2 and CSBM5. These results show consistent gains over the source model across the controlled shifts, while the relative advantage over adaptation baselines remains scenario-dependent. Appendix D.3 provides further comparisons.

Temporal shifts. Tables 1 and 2 report ogbn-arxiv and Elliptic results. On ogbn-arxiv, WaveGuard achieves the highest mean accuracy, 54.11%, which is 0.75 points above SOGA. It leads in four windows, although SOGA remains stronger on Arxiv4 and Arxiv6. On Elliptic, WaveGuard leads on both metrics in all three periods: mean Balanced Accuracy and Macro-F1 reach 75.14% and 66.70%, improving over Matcha-LAME by 6.12 and 6.63 points. Gains in both metrics indicate that the improvement extends beyond majority-class accuracy as the graph evolves.

Structural contamination and cross-lingual transfer. Figure 4 evaluates PubMed with increasing edge replacement. As replacement rises from 30% to 60%, source-model accuracy falls from 80.2% to 75.3%, while WaveGuard obtains 83.7%, 82.2%, and 82.5%. The growing improvement, from 3.5 to 7.2 points, supports its response to explicit structural contamination. Twitch provides a broader mixed-shift test: WaveGuard leads on DE→ENGB (57.46%) and DE→FR (56.88%), but

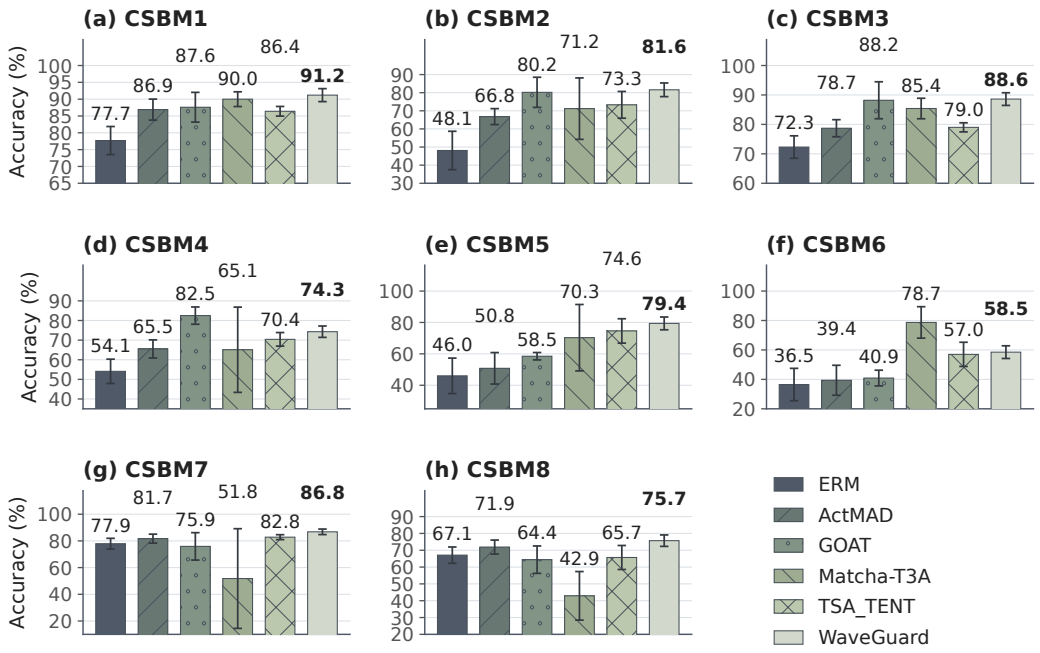

Figure 3: Target node classification accuracy (%) of representative methods on CSBM1 through CSBM8. Bars show the mean over five runs, and error bars indicate one standard deviation. Panels (a)–(d) cover structural and degree shifts, and panels (e)–(h) cover the remaining shift settings.

Table 1: Target node classification accuracy (%) across six time-shift scenarios on ogbn-arxiv. Subscripts give ± one standard deviation. The best mean is bolded and the second-best is underlined. Appendix Table G1 gives the same results in two panels.

Method	Arxiv1	Arxiv2	Arxiv3	Arxiv4	Arxiv5	Arxiv6	Avg.
ERM (Target)	$45.77_{\pm 0.42}$	$48.37_{\pm 0.51}$	$53.86_{\pm 0.47}$	$53.23_{\pm 0.55}$	$58.37_{\pm 0.62}$	$55.73_{\pm 0.58}$	52.56
ActMAD	$45.73_{\pm 0.45}$	$48.32_{\pm 0.49}$	$53.95_{\pm 0.53}$	$53.69_{\pm 0.51}$	$58.76_{\pm 0.57}$	$56.11_{\pm 0.54}$	52.76
GTrans	$45.71_{\pm 0.36}$	$47.92_{\pm 0.41}$	$53.99_{\pm 0.44}$	$53.62_{\pm 0.42}$	$58.53_{\pm 0.48}$	$56.45_{\pm 0.46}$	52.70
HomoTTT	$42.71_{\pm 0.98}$	$43.67_{\pm 1.12}$	$49.14_{\pm 1.05}$	$47.43_{\pm 1.20}$	$56.39_{\pm 1.14}$	$53.66_{\pm 1.08}$	48.83
SOGA	$42.89_{\pm 2.86}$	$49.47_{\pm 3.15}$	$53.41_{\pm 2.94}$	**57.20**$_{\pm 3.22}$	$57.68_{\pm 3.07}$	**59.46**$_{\pm 3.31}$	53.35
LAME	$44.72_{\pm 0.82}$	$49.87_{\pm 0.91}$	$54.11_{\pm 0.87}$	$53.98_{\pm 0.94}$	$58.82_{\pm 0.90}$	$54.97_{\pm 0.85}$	52.75
T3A	$43.35_{\pm 1.02}$	$45.60_{\pm 1.14}$	$51.86_{\pm 1.07}$	$50.08_{\pm 1.18}$	$55.25_{\pm 1.11}$	$52.04_{\pm 1.04}$	49.70
TENT	$45.24_{\pm 0.33}$	$47.72_{\pm 0.38}$	$53.83_{\pm 0.35}$	$53.90_{\pm 0.40}$	$58.51_{\pm 0.37}$	$56.97_{\pm 0.39}$	52.70
TSA_LAME	$45.94_{\pm 0.86}$	$50.28_{\pm 0.95}$	$54.34_{\pm 0.90}$	$54.86_{\pm 0.97}$	$58.73_{\pm 0.93}$	$54.97_{\pm 0.89}$	53.19
TSA_T3A	$44.75_{\pm 1.05}$	$47.64_{\pm 1.17}$	$53.35_{\pm 1.10}$	$53.08_{\pm 1.21}$	$56.22_{\pm 1.15}$	$53.79_{\pm 1.08}$	51.47
TSA_TENT	$45.58_{\pm 0.37}$	$47.89_{\pm 0.43}$	$54.72_{\pm 0.39}$	$54.55_{\pm 0.45}$	$58.97_{\pm 0.41}$	$56.98_{\pm 0.44}$	53.12
TTReFT	$34.06_{\pm 3.24}$	$26.91_{\pm 3.61}$	$37.30_{\pm 3.35}$	$30.50_{\pm 3.72}$	$24.19_{\pm 3.48}$	$13.15_{\pm 3.86}$	27.69
T3R	$33.36_{\pm 3.47}$	$27.81_{\pm 3.83}$	$32.50_{\pm 3.56}$	$19.13_{\pm 3.95}$	$44.65_{\pm 3.71}$	$37.16_{\pm 3.64}$	32.44
WaveGuard	**46.36**$_{\pm 0.41}$	**51.87**$_{\pm 0.46}$	**55.07**$_{\pm 0.43}$	$55.27_{\pm 0.48}$	**59.59**$_{\pm 0.44}$	$56.47_{\pm 0.47}$	**54.11**

trails the strongest method on DE→ES. Its mean of 56.34% is 0.11 points below TSA_T3A. These target-dependent results distinguish the evidence for structural adaptation from a general claim about attribute or label shifts; Table 3 reports all Twitch comparisons.

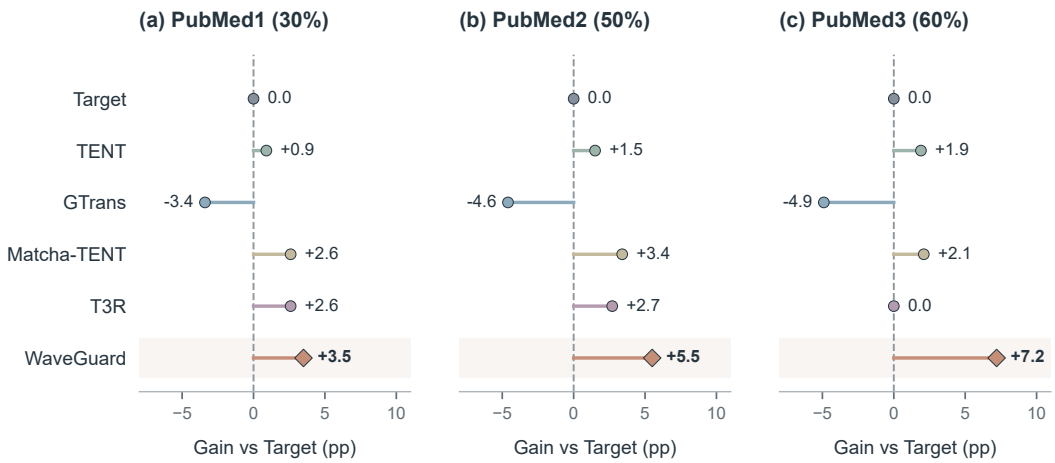

Figure 4: PubMed mean accuracy gains over Target at three edge replacement rates (percentage points). Absolute means and standard deviations appear in Appendix Figure D1.

Table 2: Mean Elliptic results (%). BA: Balanced Accuracy. Per-period means and standard deviations: Appendix Table D2.

Metric	ERM (Target)	TENT	TSA_TENT	Matcha-LAME	WaveGuard
BA	55.38	59.14	59.13	69.02	**75.14**
Macro-F1	44.93	49.17	49.07	60.07	**66.70**

Table 3: Target node classification accuracy (%) in the Twitch cross-lingual transfer scenario, including DE→ ENGB, DE→ ES, DE→ FR, and their average. Subscripts give ± one standard deviation.

Method	**DE→ENGB**	**DE→ES**	**DE→FR**	**Avg.**
ERM (Target)	$55.92_{\pm 0.31}$	$34.86_{\pm 0.42}$	$40.35_{\pm 0.36}$	43.71
ActMAD	$54.59_{\pm 0.34}$	$70.64_{\pm 0.47}$	$36.87_{\pm 0.33}$	54.03
GTrans	$56.00_{\pm 0.45}$	$35.44_{\pm 0.51}$	$41.32_{\pm 0.48}$	44.25
HomoTTT	$55.82_{\pm 0.37}$	$35.37_{\pm 0.41}$	$40.85_{\pm 0.34}$	44.01
LAME	$55.78_{\pm 0.32}$	$37.92_{\pm 0.38}$	$43.02_{\pm 0.35}$	45.57
Matcha-LAME	$47.14_{\pm 0.30}$	$67.58_{\pm 0.38}$	$53.03_{\pm 0.42}$	55.92
Matcha-T3A	$48.91_{\pm 0.26}$	$62.12_{\pm 0.29}$	$53.81_{\pm 0.36}$	54.95
Matcha-TENT	$52.28_{\pm 0.31}$	$31.63_{\pm 0.34}$	$38.33_{\pm 0.33}$	40.75
SOGA	$56.04_{\pm 0.35}$	$34.51_{\pm 0.33}$	$40.36_{\pm 0.31}$	43.64
T3A	$55.24_{\pm 0.29}$	$57.28_{\pm 0.30}$	$53.60_{\pm 0.27}$	55.37
T3R	$54.43_{\pm 0.43}$	$29.65_{\pm 0.40}$	$37.73_{\pm 0.44}$	40.60
TENT	$56.42_{\pm 0.32}$	$46.35_{\pm 0.34}$	$49.06_{\pm 0.33}$	50.61
TSA_LAME	$55.82_{\pm 0.34}$	$37.81_{\pm 0.36}$	$42.52_{\pm 0.38}$	45.38
TSA_T3A	$55.32_{\pm 0.24}$	$59.41_{\pm 0.26}$	$54.63_{\pm 0.28}$	56.45
TSA_TENT	$56.40_{\pm 0.31}$	$46.35_{\pm 0.32}$	$49.06_{\pm 0.30}$	50.60
TTReFT	$55.56_{\pm 0.50}$	$32.73_{\pm 0.58}$	$39.35_{\pm 0.62}$	42.55
WaveGuard	$57.46_{\pm 0.46}$	$54.69_{\pm 0.42}$	$56.88_{\pm 0.45}$	56.34

5.3 Selectivity and Complementarity

Selective edge attenuation. On PubMed3, post-adaptation labels separate clean homophilic, original heterophilic, and injected heterophilic edges for analysis; they do not enter score construction or weighting. The cross-run means of the within-run median retained weights are 0.956 ± 0.005, 0.444 ± 0.014, and 0.154 ± 0.005, respectively (Appendix Figure D2). The hierarchy is consistent with

selective control: original heterophilic edges remain partially active, whereas injected contamination receives the strongest attenuation. This separates the behavior of the score-guided mechanism from uniform removal of heterophilic connections.

Adaptation dynamics. A 30-step CSBM8 stress test compares the full method, each intervention alone, entropy minimization, and the source model (Appendix Figure D3). Full WaveGuard reaches $75.68\% \pm 3.56\%$, 8.57 points above the source model. Both individual interventions remain beneficial but weaker, while entropy minimization improves briefly before falling below the source model. Together with the edge-level analysis, this trajectory supports complementary control of forward messages and adaptation gradients. Appendix D.5 provides the complete interpretation.

5.4 Ablations and Hyperparameter Sensitivity

Component contributions. Figure 5 compares six ablation variants across seven scenarios: CSBM1, CSBM2, CSBM5, PubMed3, Elliptic1, Elliptic3, and Twitch1. Every variant reduces the mean performance in all seven scenarios. Removing score guidance causes the largest drop in each scenario, averaging 8.17 percentage points. Removing the score-calibrated objective, degree conditioning, or purification reduces performance by 3.67, 3.61, and 2.87 points on average; removing the anchor loss and replacing multiscale wavelets with a single scale costs 2.17 and 1.98 points. These results support the contributions of the conditional spectral reference and both intervention branches. Appendix D.6 reports the complete results.

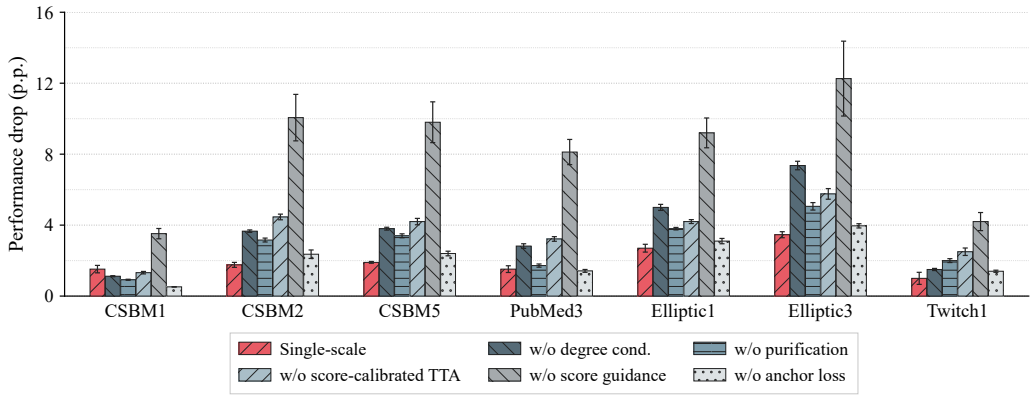

Figure 5: Performance drops relative to full WaveGuard for six ablation variants across seven scenarios. Bars show mean drops in percentage points; error bars show one standard deviation. Larger drops indicate a greater loss of performance when the corresponding component is removed or replaced.

Sensitivity. Appendix Figure D4 varies intervention-score temperature τ_r, purification strength η, and score exponent γ. Performance is stable around $\tau_r = 1$–2, $\eta = 0.7$–0.9, and $\gamma = 2$–4 on CSBM2, PubMed3, and Elliptic3. The defaults $\tau_r = 1.0$, $\eta = 0.7$, and $\gamma = 2$ lie in this shared region. Appendix D.7 discusses the individual trends.

6 Conclusion

WaveGuard coordinates message passing and unsupervised adaptation through a fixed, source-referenced node-wise intervention score. Conditional graph-wavelet statistics and frozen-model uncertainty determine where to attenuate messages and how to weight entropy, anchor, and consistency losses. Experiments support this coupling under structural and temporal shifts, while cross-lingual results remain target-dependent. The method assumes representative source summaries and transductive access to the target graph. Extending score calibration to evolving target streams, rare class–degree conditions, and directed or heterogeneous graphs remains future work.

AI USE STATEMENT

We used GPT-5.6 Sol (OpenAI) for English-language editing, figure color and layout refinement, editing plotting scripts, LaTeX typesetting, manuscript consistency checks, and reference lookup, metadata verification, and bibliography formatting.

No generative AI was used to produce synthetic research data; develop the theoretical framework; originate mathematical statements or draft, revise, or supply steps in mathematical proofs; generate or refine research hypotheses; devise or substantively assess research methods or experiments; implement the research algorithms; translate manuscript content; preprocess research datasets; or draw scientific interpretations from experimental findings. Assistance with qualitative or thematic analysis is not applicable to this study.

The authors checked AI-assisted edits against the manuscript and underlying research materials and verified that these edits preserved the reported results, mathematical meaning, and method descriptions. The authors remain responsible for the entire submission, including AI-assisted text, figures, and other materials.

REFERENCES

Wenxuan Bao, Zhichen Zeng, Zhining Liu, Hanghang Tong, and Jingrui He. Matcha: Mitigating Graph Structure Shifts with Test-Time Adaptation. In *International Conference on Learning Representations*, volume 2025, pp. 56906–56944, 2025. URL https://openreview.net/forum?id=EpgoFFUM2q.

Malik Boudiaf, Romain Mueller, Ismail Ben Ayed, and Luca Bertinetto. Parameter-Free Online Test-Time Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8344–8353, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Boudiaf_Parameter-Free_Online_Test-Time_Adaptation_CVPR_2022_paper.html.

Ruichu Cai, Fengzhu Wu, Zijian Li, Pengfei Wei, Lingling Yi, and Kun Zhang. Graph Domain Adaptation: A Generative View. *ACM Transactions on Knowledge Discovery from Data*, 18(3): 1–24, 2024. doi: 10.1145/3631712. URL https://doi.org/10.1145/3631712.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive Universal Generalized PageRank Graph Neural Network. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=n6j17fLxrP.

Tianshuo Cong, Xinlei He, Yun Shen, and Yang Zhang. Test-Time Poisoning Attacks Against Test-Time Adaptation Models. In *2024 IEEE Symposium on Security and Privacy (SP)*, pp. 1306–1324. IEEE, 2024. doi: 10.1109/sp54263.2024.00072. URL https://doi.org/10.1109/SP54263.2024.00072.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/04df4d434d481c5bb723be1b6df1ee65-Paper.pdf.

Yash Deshpande, Subhabrata Sen, Andrea Montanari, and Elchanan Mossel. Contextual Stochastic Block Models. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/08fc80de8121419136e443a70489c123-Paper.pdf.

Taesik Gong, Yewon Kim, Taeckyung Lee, Sorn Chottananurak, and Sung-Ju Lee. SoTTA: Robust Test-Time Adaptation on Noisy Data Streams. In *Advances in Neural Information Processing Systems*, volume 36, pp. 14070–14093. Curran Associates, Inc., 2023. doi: 10.52202/075280-0620. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/2da53cd1abdae59150e35f4693834f32-Abstract-Conference.html.

Kai Guo, Hongzhi Wen, Wei Jin, Yaming Guo, Jiliang Tang, and Yi Chang. Investigating Out-of-Distribution Generalization of GNNs: An Architecture Perspective. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, pp. 932–943. ACM, 2024. doi: 10.1145/3637528.3671792. URL https://doi.org/10.1145/3637528.3671792.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive Representation Learning on Large Graphs. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/5dd9db5e033da9c6fb5ba83c7a7ebea9-Paper.pdf.

David K. Hammond, Pierre Vandergheynst, and Rémi Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011. doi: 10.1016/j.acha.2010.04.005. URL https://doi.org/10.1016/j.acha.2010.04.005.

Hans Hao-Hsun Hsu, Shikun Liu, Han Zhao, and Pan Li. Structural Alignment Improves Graph Test-Time Adaptation. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 829–837. PMLR, 2026. URL https://proceedings.mlr.press/v300/hsu26a.html.

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. In *Advances in Neural Information Processing Systems*, volume 33, pp. 22118–22133. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/fb60d411a5c5b72b2e7d3527cfc84fd0-Paper.pdf.

Yusuke Iwasawa and Yutaka Matsuo. Test-Time Classifier Adjustment Module for Model-Agnostic Domain Generalization. In *Advances in Neural Information Processing Systems*, volume 34, pp. 2427–2440. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/1415fe9fea0fa1e45dddcff5682239a0-Paper.pdf.

Wei Jin, Tong Zhao, Jiayuan Ding, Yozen Liu, Jiliang Tang, and Neil Shah. Empowering Graph Representation Learning with Test-Time Graph Transformation. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Lnxl5pr018.

Mingxuan Ju, Tong Zhao, Wenhao Yu, Neil Shah, and Yanfang Ye. GraphPatcher: Mitigating Degree Bias for Graph Neural Networks via Test-time Augmentation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 55785–55801. Curran Associates, Inc., 2023. doi: 10.52202/075280-2434. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/ae9bbdcea94d808882f3535e8ca00542-Paper-Conference.pdf.

Juwon Kang, Nayeong Kim, Donghyeon Kwon, Jungseul Ok, and Suha Kwak. Leveraging Proxy of Training Data for Test-Time Adaptation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 15737–15752. PMLR, 2023. URL https://proceedings.mlr.press/v202/kang23a.html.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Haoyang Li, Xin Wang, Zeyang Zhang, Haibo Chen, Ziwei Zhang, and Wenwu Zhu. Disentangled Graph Self-supervised Learning for Out-of-Distribution Generalization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 28890–28904. PMLR, 2024a. URL https://proceedings.mlr.press/v235/li24br.html.

Haoyang Li, Xin Wang, Ziwei Zhang, and Wenwu Zhu. Out-of-Distribution Generalization on Graphs: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11):10490–10512, 2025. doi: 10.1109/tpami.2025.3593897. URL https://doi.org/10.1109/TPAMI.2025.3593897.

Xiner Li, Shurui Gui, Youzhi Luo, and Shuiwang Ji. Graph Structure Extrapolation for Out-of-Distribution Generalization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 27846–27874. PMLR, 2024b. URL https://proceedings.mlr.press/v235/li24y.html.

Jian Liang, Ran He, and Tieniu Tan. A Comprehensive Survey on Test-Time Adaptation Under Distribution Shifts. *International Journal of Computer Vision*, 133(1):31–64, 2025. doi: 10.1007/s11263-024-02181-w. URL https://doi.org/10.1007/s11263-024-02181-w.

Meihan Liu, Zhen Zhang, Jiachen Tang, Jiajun Bu, Bingsheng He, and Sheng Zhou. Revisiting, Benchmarking and Understanding Unsupervised Graph Domain Adaptation. In *Advances in Neural Information Processing Systems*, volume 37, pp. 89408–89436. Curran Associates, Inc., 2024a. doi: 10.52202/079017-2838. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/a2c548b2ed3141400d86a235eaa6fad0-Paper-Datasets_and_Benchmarks_Track.pdf.

Shikun Liu, Tianchun Li, Yongbin Feng, Nhan Tran, Han Zhao, Qiang Qiu, and Pan Li. Structural Re-weighting Improves Graph Domain Adaptation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 21778–21793. PMLR, 2023. URL https://proceedings.mlr.press/v202/liu23u.html.

Shikun Liu, Deyu Zou, Han Zhao, and Pan Li. Pairwise Alignment Improves Graph Domain Adaptation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 32552–32575. PMLR, 2024b. URL https://proceedings.mlr.press/v235/liu24ci.html.

Sitao Luan, Chenqing Hua, Minkai Xu, Qincheng Lu, Jiaqi Zhu, Xiao-Wen Chang, Jie Fu, Jure Leskovec, and Doina Precup. When Do Graph Neural Networks Help with Node Classification? Investigating the Homophily Principle on Node Distinguishability. In *Advances in Neural Information Processing Systems*, volume 36, pp. 28748–28760. Curran Associates, Inc., 2023. doi: 10.52202/075280-1251. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/5ba11de4c74548071899cf41dec078bf-Paper-Conference.pdf.

Junyu Luo, Yiyang Gu, Xiao Luo, Wei Ju, Zhiping Xiao, Yusheng Zhao, Jingyang Yuan, and Ming Zhang. GALA: Graph Diffusion-Based Alignment With Jigsaw for Source-Free Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9038–9051, 2024a. doi: 10.1109/tpami.2024.3416372. URL https://doi.org/10.1109/tpami.2024.3416372.

Junyu Luo, Zhiping Xiao, Yifan Wang, Xiao Luo, Jingyang Yuan, Wei Ju, Langechuan Liu, and Ming Zhang. Rank and Align: Towards Effective Source-free Graph Domain Adaptation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pp. 4706–4714. International Joint Conferences on Artificial Intelligence Organization, 2024b. doi: 10.24963/ijcai.2024/520. URL https://doi.org/10.24963/ijcai.2024/520.

Jing Ma. A survey of out-of-distribution generalization for graph machine learning from a causal view. *AI Magazine*, 45(4):537–548, 2024a. doi: 10.1002/aaai.12202. URL https://doi.org/10.1002/aaai.12202.

Jing Ma. Improved Self-Training for Test-Time Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23701–23710, 2024b. URL https://openaccess.thecvf.com/content/CVPR2024/html/Ma_Improved_Self-Training_for_Test-Time_Adaptation_CVPR_2024_paper.html.

Haitao Mao, Zhikai Chen, Wei Jin, Haoyu Han, Yao Ma, Tong Zhao, Neil Shah, and Jiliang Tang. Demystifying Structural Disparity in Graph Neural Networks: Can One Size Fit All? In *Advances in Neural Information Processing Systems*, volume 36, pp. 37013–37067. Curran Associates, Inc., 2023. doi: 10.52202/075280-1610. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/74f1edadbdf495e7258ee8db7b1d3acd-Paper-Conference.pdf.

Haitao Mao, Lun Du, Yujia Zheng, Qiang Fu, Zelin Li, Xu Chen, Shi Han, and Dongmei Zhang. Source Free Graph Unsupervised Domain Adaptation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, WSDM '24, pp. 520–528. ACM, 2024. doi: 10.1145/3616855.3635802. URL https://doi.org/10.1145/3616855.3635802.

Muhammad Jehanzeb Mirza, Pol Jané Soneira, Wei Lin, Mateusz Kozinski, Horst Possegger, and Horst Bischof. ActMAD: Activation Matching To Align Distributions for Test-Time-Training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24152–24161, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Mirza_ActMAD_Activation_Matching_To_Align_Distributions_for_Test-Time-Training_CVPR_2023_paper.html.

Oleg Platonov, Denis Kuznedelev, Artem Babenko, and Liudmila Prokhorenkova. Characterizing Graph Datasets for Node Classification: Homophily-Heterophily Dichotomy and Beyond. In *Advances in Neural Information Processing Systems*, volume 36, pp. 523–548. Curran Associates, Inc., 2023. doi: 10.52202/075280-0025. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/01b681025fdbda8e935a66cc5bb6e9de-Paper-Conference.pdf.

Ziyue Qiao, Qianyi Cai, Hao Dong, Jiawei Gu, Pengyang Wang, Meng Xiao, Xiao Luo, and Hui Xiong. GCAL: Adapting Graph Models to Evolving Domain Shifts. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 50113–50129. PMLR, 2025. URL https://proceedings.mlr.press/v267/qiao25a.html.

Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-Scale attributed node embedding. *Journal of Complex Networks*, 9(2):cnab014, 2021. doi: 10.1093/comnet/cnab014. URL https://doi.org/10.1093/comnet/cnab014.

Yongduo Sui, Qitian Wu, Jiancan Wu, Qing Cui, Longfei Li, Jun Zhou, Xiang Wang, and Xiangnan He. Unleashing the Power of Graph Data Augmentation on Covariate Distribution Shift. In *Advances in Neural Information Processing Systems*, volume 36, pp. 18109–18131. Curran Associates, Inc., 2023. doi: 10.52202/075280-0796. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/3a33ddacb2798fc7d83b8334d552e05a-Paper-Conference.pdf.

Haopeng Sun, Lumin Xu, Sheng Jin, Ping Luo, Chen Qian, and Wentao Liu. PROGRAM: PROtotype GRAph Model based Pseudo-Label Learning for Test-Time Adaptation. In *International Conference on Learning Representations*, volume 2024, pp. 34298–34328, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/92369a01fbe8046a093746389b2c413e-Abstract-Conference.html.

Huy Truong, Alexander Lazovik, and Victoria Degeler. T3R: Deeper Test-Time Adaptation for Graph Neural Networks via Gradient Rotation, 2026. URL https://arxiv.org/abs/2606.30011.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully Test-Time Adaptation by Entropy Minimization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=uXl3bZLkr3c.

Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I. Weidele, Claudio Bellei, Tom Robinson, and Charles E. Leiserson. Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics. In *KDD Workshop on Anomaly Detection in Finance*, 2019. URL https://arxiv.org/abs/1908.02591.

Qitian Wu, Fan Nie, Chenxiao Yang, and Junchi Yan. Learning Divergence Fields for Shift-Robust Graph Representations. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53776–53792. PMLR, 2024. URL https://proceedings.mlr.press/v235/wu24v.html.

Charles Xu, Laney Goldman, Valentina Guo, Benjamin Hollander-Bodie, Maedee Trank-Greene, Ian Adelstein, Edward De Brouwer, Rex Ying, Smita Krishnaswamy, and Michael Perlmutter.

BLIS-Net: Classifying and Analyzing Signals on Graphs. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238, pp. 4537–4545. PMLR, 2024a. URL https://proceedings.mlr.press/v238/xu24c.html.

Haobo Xu, Yuchen Yan, Dingsu Wang, Zhe Xu, Zhichen Zeng, Tarek F. Abdelzaher, Jiawei Han, and Hanghang Tong. SLOG: An Inductive Spectral Graph Neural Network Beyond Polynomial Filter. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 55348–55370. PMLR, 2024b. URL https://proceedings.mlr.press/v235/xu24aa.html.

Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. Revisiting Semi-Supervised Learning with Graph Embeddings. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 40–48. PMLR, 2016. URL https://proceedings.mlr.press/v48/yanga16.html.

Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. Graph Domain Adaptation via Theory-Grounded Spectral Regularization. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OysfLgrk8mk.

Junchi Yu, Jian Liang, and Ran He. Mind the Label Shift of Augmentation-Based Graph OOD Generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11620–11630, 2023.

Longhui Yuan, Binhui Xie, and Shuang Li. Robust Test-Time Adaptation in Dynamic Scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15922–15932, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Yuan_Robust_Test-Time_Adaptation_in_Dynamic_Scenarios_CVPR_2023_paper.html.

Haoxiang Zhang, Zhuofeng Li, Qiannan Zhang, Ziyi Kou, Juncheng Li, and Shichao Pei. Avoiding Structural Pitfalls: Self-Supervised Low-Rank Feature Tuning for Graph Test-Time Adaptation. *Transactions on Machine Learning Research*, 2025. URL https://openreview.net/forum?id=yiS6q42LLt.

Jiaxin Zhang, Yiqi Wang, Xihong Yang, and En Zhu. A Fully Test-time Training Framework for Semi-supervised Node Classification on Out-of-Distribution Graphs. *ACM Transactions on Knowledge Discovery from Data*, 18(7):1–19, 2024a. doi: 10.1145/3649507. URL https://doi.org/10.1145/3649507.

Jiaxin Zhang, Yiqi Wang, Siwei Wang, Xihong Yang, Yu Shi, Xinwang Liu, and En Zhu. Beyond Parameter Finetuning: Test-Time Representation Refinement for Node Classification, 2026. URL https://arxiv.org/abs/2601.21615.

Zhen Zhang, Meihan Liu, Anhui Wang, Hongyang Chen, Zhao Li, Jiajun Bu, and Bingsheng He. Collaborate to Adapt: Source-Free Graph Domain Adaptation via Bi-directional Adaptation. In *Proceedings of the ACM Web Conference 2024*, WWW '24, pp. 664–675. ACM, 2024b. doi: 10.1145/3589334.3645507. URL https://doi.org/10.1145/3589334.3645507.

Hao Zhao, Yuejiang Liu, Alexandre Alahi, and Tao Lin. On Pitfalls of Test-Time Adaptation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 42058–42080. PMLR, 2023. URL https://proceedings.mlr.press/v202/zhao23d.html.

Yusheng Zhao, Qixin Zhang, Xiao Luo, Junyu Luo, Wei Ju, Zhiping Xiao, and Ming Zhang. Test-time Adaptation on Graphs via Adaptive Subgraph-based Selection and Regularized Prototypes. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 78003–78022. PMLR, 2025. URL https://proceedings.mlr.press/v267/zhao25ai.html.

Zhi Zhou, Lan-Zhe Guo, Lin-Han Jia, Dingchu Zhang, and Yu-Feng Li. ODS: Test-Time Adaptation in the Presence of Open-World Data Shift. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 42574–42588. PMLR, 2023. URL https://proceedings.mlr.press/v202/zhou23e.html.

A Problem Setup and Preliminaries

A.1 Graph Setting and Distribution-Shift Definitions

Graph node classification. We study transductive, source-free GTTA on an attributed graph

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{A}). \tag{A.1}$$

where $\mathcal{V}$ and $\mathcal{E}$ are the node and edge sets, $N = |\mathcal{V}|$, $\mathbf{X} \in \mathbb{R}^{N \times d}$ contains node features, and $\mathbf{A} \in \{0,1\}^{N \times N}$ is the adjacency matrix. Labels are $\mathbf{Y} \in \{0,1\}^{N \times C}$ for C classes. Node v_i has neighbors $\mathcal{N}(v_i)$, degree $d_i = |\mathcal{N}(v_i)|$, and pretrained prediction $\mathbf{p}_i = [f_{\theta_0}(\mathbf{X}, \mathbf{A})]_i$.

The source model f_{θ_0} is trained on labeled $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s, \mathbf{X}_s, \mathbf{A}_s)$ with labels $\mathbf{Y}_s$, then adapted to an unlabeled target graph $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \mathbf{X}_t, \mathbf{A}_t)$. Unlike batch-wise image TTA, transductive GTTA exposes the entire target graph. Target labels $\mathbf{Y}_t$ serve only conceptual definitions and final evaluation. Deployment retains model parameters and compact source statistics, but no source features, adjacency, labels, or per-node representations. This setting is also termed source-free unsupervised graph domain adaptation (Liang et al., 2025; Jin et al., 2023; Ju et al., 2023; Mao et al., 2024; Zhang et al., 2024b; Bao et al., 2025; Zhao et al., 2025; Qiao et al., 2025; Zhang et al., 2025; Hsu et al., 2026; Luo et al., 2024a;b; Zhao et al., 2023; Gong et al., 2023; Yuan et al., 2023; Zhou et al., 2023; Kang et al., 2023; Sun et al., 2024; Ma, 2024b; Cong et al., 2024; Mirza et al., 2023).

All input graphs are made undirected and duplicate edges merged, so $\mathbf{A} = \mathbf{A}^\top$. For support edge $\{i, j\}$, $j \to i$ denotes an aggregation direction, not a directed dataset edge. Adjacency determines which node features are mixed during message passing.

Attribute shift. The class-conditional feature distribution changes when

$$P_s(\mathbf{x} \mid y) \neq P_t(\mathbf{x} \mid y), \tag{A.2}$$

Thus, nodes of the same class can present different feature patterns across domains.

Structural shift. Connectivity patterns change when the joint distribution of adjacency and labels satisfies

$$P_s(\mathbf{A}, \mathbf{Y}) \neq P_t(\mathbf{A}, \mathbf{Y}), \tag{A.3}$$

This distributional definition does not label individual edges as anomalous. We consider degree and homophily shifts, local rewiring, and temporal evolution (Mao et al., 2023; Platonov et al., 2023; Luan et al., 2023; Li et al., 2025; Liu et al., 2023; 2024a;b; You et al., 2023).

Degree shift. Changes in mean degree, low-degree fraction, or tail, arising from growth, activity, or sampling, alter neighborhood size, self/neighbor balance, and multilayer propagation reach.

Homophily shift. For $d_i > 0$, node homophily h_i and graph-average homophily are

$$h(\mathcal{G}) = \frac{1}{N} \sum_{i=1}^{N} h_i, \qquad h_i = \frac{|\{v_j \in \mathcal{N}(v_i) : y_j = y_i\}|}{d_i}. \tag{A.4}$$

A homophily shift satisfies $h(\mathcal{G}_s) \neq h(\mathcal{G}_t)$. Lower homophily increases semantically inconsistent mixing. Equation (A.4) uses labels only for shift definition and offline evaluation, never for test-time adaptation.

Other structural changes. Local rewiring can change individual neighborhoods while approximately preserving global degree and homophily; temporal evolution can add, delete, or rewire edges. WaveGuard therefore uses node-wise neighborhood and multiscale spectral responses rather than one global statistic. Attribute and label shifts, $P_s(y) \neq P_t(y)$, may coexist but are not its primary calibration targets.

A.2 Source-Free GTTA Objective and Information Budget

Task objective. Using the source and target graphs defined in Appendix A.1, GTTA adapts pre-trained parameters θ_0 without observing $\mathbf{Y}_t$.

The ideal supervised target risk is

$$\mathcal{R}_t(\theta, \mathbf{W}^t) = \frac{1}{N_t} \sum_{i=1}^{N_t} \ell\left(\left[f_\theta\left(\mathbf{X}_t, \mathbf{W}^t\right)\right]_i, y_i^t\right), \tag{A.5}$$

where ℓ is the classification loss and $\mathbf{W}^t$ the directed message-weight matrix. Since $\mathbf{Y}_t$ is unavailable, this risk cannot guide updates, model selection, or stopping. GTTA instead optimizes an unlabeled surrogate from predictions, augmented views, or structural statistics (Liang et al., 2025; Jin et al., 2023; Mao et al., 2024):

$$(\theta^*, \mathbf{W}^{t*}) = \underset{\theta \in \Theta_{\text{adapt}},\ \mathbf{W}^t \in \mathcal{W}_t}{\operatorname{argmin}} \mathcal{L}_{\text{TTA}}\left(f_\theta; \mathbf{X}_t, \mathbf{W}^t, \theta_0, \mathcal{M}_s\right). \tag{A.6}$$

Here $\mathcal{L}_{\text{TTA}}$ is the unlabeled loss, Θ_{adapt} the updatable parameter set, $\mathcal{M}_s$ the retained source summary, and $\mathcal{W}_t$ the allowed message-weight matrices. Parameter-only methods fix $\mathbf{W}^t = \mathbf{A}_t$; graph-only methods fix $\theta = \theta_0$. WaveGuard controls both propagation and parameter updates.

Update scope. Before deployment, select Θ_{adapt} from all parameters or normalization, classifier, and other modules. Restricted updates reduce cost but may leave aggregation-induced corruption unresolved. WaveGuard supports any subset, emphasizing lower-score nodes in entropy and consistency while intervening more strongly on higher-score nodes.

Message weights. The original adjacency remains undirected and symmetric. WaveGuard assigns directed message weights on its existing support:

$$\mathcal{W}_t = \left\{\mathbf{W} \in [0,1]^{N_t \times N_t} \;\middle|\; \operatorname{supp}(\mathbf{W}) \subseteq \operatorname{supp}(\mathbf{A}_t)\right\}. \tag{A.7}$$

No symmetry constraint is imposed on $\mathbf{W}$; opposite directions may have different weights while the original support remains unchanged. Main Section 4 defines the scores and directional message weights.

Source information budget. Deployment retains θ_0 and a precomputed aggregate summary

$$\mathcal{M}_s = \text{Summary}\left(\mathcal{G}_s, f_{\theta_0}\right). \tag{A.8}$$

The summary depends on the numbers of classes, degree buckets, and spectral scales, rather than source-node count N_s. It excludes source-node features, adjacency, labels, and per-node representations. WaveGuard stores conditional spectral statistics, scale weights, and deviation thresholds (Appendix B.1–B.2), allowing a source reference without source-data storage or replay.

A.3 Normalized Laplacian and Multiscale Graph Wavelets

Normalized Laplacian. For an undirected graph, let $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_N)$. Define

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}. \tag{A.9}$$

For isolated nodes, the corresponding entry of $\mathbf{D}^{-1/2}$ is zero. Degree normalization reduces the direct influence of absolute degree; the eigenvalues lie in $[0, 2]$.

The symmetric positive-semidefinite matrix has eigendecomposition

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}, \tag{A.10}$$

where the orthogonal eigenvectors $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]$ form the graph Fourier basis and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$ contains ascending frequencies. For $\mathbf{H} \in \mathbb{R}^{N \times d_h}$, the Fourier transform is $\widehat{\mathbf{H}} = \mathbf{U}^\mathsf{T}\mathbf{H}$, with inverse $\mathbf{H} = \mathbf{U}\widehat{\mathbf{H}}$.

Signal variation is measured by the Dirichlet energy:

$$\operatorname{tr}\left(\mathbf{H}^{\mathsf{T}}\mathbf{L}\mathbf{H}\right) = \frac{1}{2}\sum_{i,j} A_{\mathrm{ij}} \parallel \frac{\mathbf{h}_i}{\sqrt{d_i}} - \frac{\mathbf{h}_j}{\sqrt{d_j}} \parallel_2^2 . \tag{A.11}$$

Similar degree-normalized neighbors have low energy; local differences contribute higher frequencies. Structural shifts redistribute this energy by changing neighbors and normalization, even with unchanged attributes.

Multiscale graph wavelets. To resolve structural responses across frequency bands, apply mother kernel g at scale s (Defferrard et al., 2016; Hammond et al., 2011):

$$\mathbf{\Psi}_s = g_s(\mathbf{L}) = \mathbf{U}g(s\mathbf{\Lambda})\mathbf{U}^{\mathsf{T}}. \tag{A.12}$$

where

$$g(s\mathbf{\Lambda}) = \operatorname{diag}\left(g(s\lambda_1), \ldots, g(s\lambda_N)\right) \tag{A.13}$$

sets the frequency response. Smaller scales emphasize higher-frequency, localized changes; larger scales cover lower frequencies and broader propagation. The scale set $\mathcal{S} = \{s_1, \ldots, s_K\}$ captures both.

For frozen-model hidden representations $\mathbf{H}$, the response is

$$\mathbf{Z}_s = \mathbf{\Psi}_s\mathbf{H}. \tag{A.14}$$

The ith row $\mathbf{z}_{i,s}$ is the response centered on node v_i. Its local energy is

$$e_{i,s} = \parallel \mathbf{z}_{i,s} \parallel_2^2, \tag{A.15}$$

A vector composed of multiple scales

$$\mathbf{e}_i = [e_{i,s_1}, e_{i,s_2}, \ldots, e_{i,s_K}] \tag{A.16}$$

retains node-level spectral differences that a graph-average energy would obscure.

Sparse approximation. Mapping the spectrum from $[0, 2]$ to $[-1, 1]$ permits an order-R Chebyshev approximation without explicit eigendecomposition:

$$\mathbf{\Psi}_s\mathbf{H} \approx \sum_{r=0}^{R} \alpha_r^{(s)} T_r(\widetilde{\mathbf{L}})\mathbf{H}, \tag{A.17}$$

Here $\widetilde{\mathbf{L}} = \mathbf{L} - \mathbf{I}$, T_r is the order-r Chebyshev polynomial, and $\alpha_r^{(s)}$ its scale-specific coefficient. Sparse matrix recurrences cost approximately $\mathcal{O}(R|\mathcal{E}|d_h)$ per scale. Appendix B.1–B.2 specifies the kernel, representation layer, scales, and order.

A.4 Unlabeled Test-Time Objectives under Structural Shift

Without target labels, entropy minimization and consistency learning supply adaptation signals from predictions and perturbed views (Liang et al., 2025; Jin et al., 2023; Mao et al., 2024). Their usefulness depends on whether those signals align with classification correctness.

Entropy minimization. For target prediction

$$\mathbf{p}_i = [p_{i,1}, p_{i,2}, \ldots, p_{i,C}], \qquad \sum_{c=1}^{C} p_{i,c} = 1. \tag{A.18}$$

Its prediction entropy is defined as

$$\mathcal{H}(\mathbf{p}_i) = -\sum_{c=1}^{C} p_{i,c} \log\left(p_{i,c} + \epsilon\right), \tag{A.19}$$

where $\epsilon > 0$ is a constant used to ensure numerical stability. The standard entropy minimization objective is (Liang et al., 2025; Zhao et al., 2023; Gong et al., 2023; Yuan et al., 2023; Zhou et al., 2023; Kang et al., 2023; Sun et al., 2024; Ma, 2024b)

$$\mathcal{L}_{\text{ent}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \mathcal{H}(\mathbf{p}_i). \tag{A.20}$$

With reliable initial predictions and low-density separation, entropy minimization can move boundaries away from dense target regions, especially under weak structural shift. Confident errors from corrupted neighborhoods are instead reinforced and can spread through shared parameters.

Consistency learning. Independent dropout, feature noise, or other semantics-preserving perturbations $\mathcal{T}_{\xi_1}, \mathcal{T}_{\xi_2}$ generate predictions

$$\begin{aligned} \mathbf{p}_i^{(1)} &= \left[f_\theta\left(\mathcal{T}_{\xi_1}(\mathbf{X}_t, \mathbf{A}_t)\right)\right]_i, \\ \mathbf{p}_i^{(2)} &= \left[f_\theta\left(\mathcal{T}_{\xi_2}(\mathbf{X}_t, \mathbf{A}_t)\right)\right]_i. \end{aligned} \tag{A.21}$$

The corresponding consistency objective is defined as

$$\mathcal{L}_{\text{cons}} = \frac{1}{N_t} \sum_{i=1}^{N_t} d\left(\mathbf{p}_i^{(1)}, \mathbf{p}_i^{(2)}\right). \tag{A.22}$$

where d may be symmetric KL, Jensen–Shannon divergence, or another distributional distance. The objective reduces prediction variation under perturbations.

Both views can inherit the same corrupted neighborhood and agree on a wrong class. Strong structural augmentation can also compound the target shift. Consistency suppresses random variation around reliable predictions but cannot alone correct systematic neighborhood bias.

B Complete WaveGuard Formulation

B.1 Class–Degree–Conditioned Spectral Memory

Extract the frozen source model's final hidden representation before its classifier:

$$\mathbf{H} = f_{\theta_0}^{\text{enc}}(\mathbf{X}, \mathbf{A}) = [\mathbf{h}_1, \ldots, \mathbf{h}_N]^\top \in \mathbb{R}^{N \times d_h}. \tag{B.1}$$

Here d_h is the hidden dimension. These representations retain neighborhood information before class-probability compression. With the normalized Laplacian in Equation (A.9), use the high-pass wavelet

$$g_{s_k}(\lambda) = \lambda \exp(-s_k \lambda), \qquad \mathbf{Z}_k = g_{s_k}(\mathbf{L})\mathbf{H} \approx \sum_{r=0}^{R} a_{k,r} T_r(\mathbf{L})\mathbf{H}. \tag{B.2}$$

Here $\mathbf{Z}_k$ is the scale response, T_r a Chebyshev polynomial, and $a_{k,r}$ its coefficient. Sparse recurrences avoid eigendecomposition. We use $\mathcal{S} = \{0.05, 0.1, 0.2, 0.5, 1.0\}$ and $R = 5$, covering localized changes and broader propagation.

The node-wise energy and log-energy signature are

$$
\begin{aligned}
E_i^{(k)} &= \| \mathbf{Z}_k[i,:] \|_2^2, \\
e_i^{(k)} &= \log\left(E_i^{(k)} + \epsilon\right), \\
\mathbf{e}_i &= \left[e_i^{(1)}, \ldots, e_i^{(K)}\right]^\top.
\end{aligned}
\tag{B.3}
$$

With numerical stabilizer $\epsilon > 0$, the logarithm moderates differences in graph size, hidden dimension, and response amplitude across scales.

A global reference conflates shift with normal class and degree heterogeneity. Set $u_i = \log(1 + \deg(i))$, partition source log-degrees into B equal-frequency buckets with centers m_b, and define soft memberships

$$
\begin{gathered}
\alpha_{i,b} = \frac{\exp\left[-(u_i - m_b)^2/(2\tau_d^2)\right]}{\sum_{b'=1}^{B} \exp\left[-(u_i - m_{b'})^2/(2\tau_d^2)\right]}, \\
b = 1, \ldots, B.
\end{gathered}
\tag{B.4}
$$

The bandwidth τ_d smooths adjacent buckets, preserving degree-dependent responses without discontinuous references at bucket boundaries.

Source labels give the conditional energy moments for class c, bucket b, and scale k:

$$
\begin{aligned}
\mu_{c,b,k} &= \frac{\sum_{i:y_i=c} \alpha_{i,b} e_i^{(k)}}{\sum_{i:y_i=c} \alpha_{i,b} + \epsilon}, \\
\sigma_{c,b,k}^2 &= \frac{\sum_{i:y_i=c} \alpha_{i,b} \left(e_i^{(k)} - \mu_{c,b,k}\right)^2}{\sum_{i:y_i=c} \alpha_{i,b} + \epsilon}.
\end{aligned}
\tag{B.5}
$$

When there are insufficient valid samples for a given class-degree condition, the statistics fall back sequentially to class-level and global estimates to reduce the estimation variance in long-tail intervals.

Controlled source-edge dropout with frozen parameters measures structural sensitivity and class discriminability:

$$
\begin{aligned}
r_{c,b,k}^{\text{sens}} &= \frac{1}{|\mathcal{B}_{c,b}|} \sum_{i \in \mathcal{B}_{c,b}} \frac{\left|e_{i,clean}^{(k)} - e_{i,corr}^{(k)}\right|}{\sigma_{c,b,k} + \epsilon}, \\
r_{c,b,k}^{\text{disc}} &= \frac{|\mu_{c,b,k} - \mu_{\neg c,b,k}|}{\sigma_{c,b,k} + \sigma_{\neg c,b,k} + \epsilon}.
\end{aligned}
\tag{B.6}
$$

Here $\mathcal{B}_{c,b}$ contains source nodes of class c and bucket b; $\mu_{\neg c,b,k}, \sigma_{\neg c,b,k}$ summarize other classes at the same bucket and scale. Combine both criteria into normalized scale weights

$$
w_{c,b,k} = \frac{r_{c,b,k}^{\text{sens}} \left(r_{c,b,k}^{\text{disc}}\right)^p}{\sum_{\ell=1}^{K} r_{c,b,\ell}^{\text{sens}} \left(r_{c,b,\ell}^{\text{disc}}\right)^p + \epsilon}.
\tag{B.7}
$$

The exponent p controls class discriminability. Define $q_{c,b}$ as the 95th percentile of conditional deviations on the unperturbed source graph. Retain only

$$
\mathcal{M} = \left\{m_b, \mu_{c,b,k}, \sigma_{c,b,k}, w_{c,b,k}, q_{c,b}\right\}.
\tag{B.8}
$$

These aggregate statistics exclude the source graph and per-node representations; storage is independent of source-node count.

B.2 Target Spectral Deviation and Fixed Intervention Score

To avoid hard pseudo-label errors in conditional references, use the frozen model's full predictive distribution on the original target graph:

$$\mathbf{H}^t = f_{\theta_0}^{\text{enc}}(\mathbf{X}^t, \mathbf{A}^t), \qquad \mathbf{p}_i^0 = \text{softmax}\left(f_{\theta_0}(\mathbf{X}^t, \mathbf{A}^t)_i\right). \tag{B.9}$$

Compute target energies $\mathbf{e}_i^t$ from $\mathbf{H}^t$ and memberships $\alpha_{i,b}^t$ from target degrees. The conditional deviation is

$$D_i(c,b) = \sum_{k=1}^{K} w_{c,b,k} \left(\frac{e_{i,t}^{(k)} - \mu_{c,b,k}}{\sigma_{c,b,k} + \epsilon} \right)^2. \tag{B.10}$$

Next, soft marginalization is applied over the degree and class conditions:

$$D_i(c) = \sum_{b=1}^{B} \alpha_{i,b}^t D_i(c,b), \qquad D_i = \sum_{c=1}^{C} p_{i,c}^0 D_i(c). \tag{B.11}$$

Soft degree and class marginalization yield a continuous conditional deviation without hard bin or pseudo-label assignments.

Complement the deviation with normalized predictive entropy

$$\mathcal{H}(\mathbf{p}_i^0) = -\frac{1}{\log C} \sum_{c=1}^{C} p_{i,c}^0 \log\left(p_{i,c}^0 + \epsilon\right), \tag{B.12}$$

and calculate the node conditional thresholds based on the same class probabilities and soft degree assignment

$$q_i = \sum_{c=1}^{C} \sum_{b=1}^{B} p_{i,c}^0 \alpha_{i,b}^t q_{c,b}. \tag{B.13}$$

Combine deviation, uncertainty, and the source threshold into the intervention score:

$$\rho_i = \text{sigmoid}\left(\frac{D_i + \beta \mathcal{H}(\mathbf{p}_i^0) - q_i}{\tau_r} \right), \qquad 0 < \rho_i < 1. \tag{B.14}$$

Here β weights uncertainty and τ_r controls gating smoothness. Larger deviation or uncertainty increases ρ_i, strengthening intervention.

B.3 Direction-Aware Soft Topology Purification

For direction $j \to i$ on undirected support $\{i, j\}$, compute cosine similarity from frozen representations on the original target graph:

$$s_{j\to i} = \frac{(\mathbf{h}_j^t)^\top \mathbf{h}_i^t}{\| \mathbf{h}_j^t \|_2 \| \mathbf{h}_i^t \|_2 + \epsilon}. \tag{B.15}$$

A receiver-specific threshold accommodates local density and similarity differences. Define

$$\ell_i = \text{Quantile}_\eta \left\{ s_{j\to i} : j \in \mathcal{N}(i) \right\}, \qquad m_{j\to i} = \text{sigmoid}\left(\frac{\ell_i - s_{j\to i}}{\tau_e} \right). \tag{B.16}$$

Here $\eta \in [0,1]$ controls the local low-similarity fraction prioritized for suppression, and $\tau_e > 0$ its smoothness. Similarities below ℓ_i yield larger mismatch $m_{j\to i}$.

Because legitimate heterophilic edges can also have low similarity, combine mismatch with sender and receiver scores:

$$\begin{aligned} r_{j\to i} &= 1-(1-\rho_j)\,(1-\rho_i m_{j\to i})\,, \\ \omega_{j\to i} &= 1-(1-\omega_{\min})r_{j\to i}, \\ W^t_{ji} &= A^t_{\mathrm{ji}}\omega_{j\to i}. \end{aligned} \tag{B.17}$$

High sender scores attenuate outgoing messages. For low-score senders, attenuation increases when the receiver has both a high score and strong local mismatch. Receiver-specific thresholds generally yield $\omega_{j\to i} \neq \omega_{i\to j}$ on the same undirected support.

Self-loops are preserved and $\omega_{\min} > 0$ prevents hard edge removal under score misspecification. Normalize the directed message weights $\mathbf{W}^t$ for propagation and hold them fixed throughout adaptation; the original adjacency $\mathbf{A}^t$ remains symmetric.

B.4 Score-Calibrated Test-Time Objective

The current model predicts on the purified graph, while the frozen model supplies fixed predictions on both graphs:

$$\begin{aligned} \mathbf{p}_i^{\theta} &= \mathrm{softmax}\left(f_\theta(\mathbf{X}^t,\mathbf{W}^t)_i\right), \\ \mathbf{p}_i^{\mathrm{orig}} &= \mathrm{sg}\left[\mathrm{softmax}\left(f_{\theta_0}(\mathbf{X}^t,\mathbf{A}^t)_i\right)\right], \\ \mathbf{p}_i^{\mathrm{pur}} &= \mathrm{sg}\left[\mathrm{softmax}\left(f_{\theta_0}(\mathbf{X}^t,\mathbf{W}^t)_i\right)\right]. \end{aligned} \tag{B.18}$$

The operator sg stops gradients. The original prediction measures purification sensitivity; $\mathbf{p}_i^{\mathrm{pur}}$ supplies the stability anchor. Their divergence and purified-prediction uncertainty determine reliability:

$$\begin{aligned} \delta_i &= \mathrm{JSD}\left(\mathbf{p}_i^{\mathrm{orig}}, \mathbf{p}_i^{\mathrm{pur}}\right), \\ r_i^{\mathrm{anc}} &= \exp\left(-\frac{\delta_i}{\tau_a}\right)\left[1-\mathcal{H}\left(\mathbf{p}_i^{\mathrm{pur}}\right)\right]. \end{aligned} \tag{B.19}$$

JSD denotes Jensen–Shannon divergence, $\tau_a > 0$ controls discrepancy decay, and $\mathcal{H}$ is the normalized entropy in Equation (B.12). Disagreement or high purified-prediction entropy reduces anchor reliability.

The node weights for the entropy term and the anchor term are defined as

$$w_i^{\mathrm{ent}} = (1-\rho_i)^{\gamma}, \qquad w_i^{\mathrm{anc}} = \rho_i^{\gamma} r_i^{\mathrm{anc}}. \tag{B.20}$$

The exponent $\gamma > 0$ controls score separation. The losses are

$$\begin{aligned} \mathcal{L}_{\mathrm{ent}} &= \frac{1}{N_t}\sum_{i=1}^{N_t} w_i^{\mathrm{ent}}\mathcal{H}(\mathbf{p}_i^{\theta}), \\ \mathcal{L}_{\mathrm{anc}} &= \frac{1}{N_t}\sum_{i=1}^{N_t} w_i^{\mathrm{anc}}\,\mathrm{KL}\left(\mathbf{p}_i^{\mathrm{pur}} \parallel \mathbf{p}_i^{\theta}\right). \end{aligned} \tag{B.21}$$

Lower-score nodes receive larger entropy weights; higher-score nodes receive larger anchor weights only when the anchor is reliable. Sharing the purified graph makes the anchor KL term primarily constrain parameter drift.

Two independent dropout passes on $\mathbf{W}^t$ yield $\mathbf{p}_i^{(1)}, \mathbf{p}_i^{(2)}$. For each node, select the view with larger maximum class probability as teacher $\mathbf{p}_i^{\mathrm{hi}}$, and the other as student $\mathbf{p}_i^{\mathrm{lo}}$:

$$\mathcal{L}_{\mathrm{cons}} = \frac{1}{N_t}\sum_{i=1}^{N_t} w_i^{\mathrm{ent}}\,\mathrm{KL}\left(\mathrm{sg}[\mathbf{p}_i^{\mathrm{hi}}] \parallel \mathbf{p}_i^{\mathrm{lo}}\right). \tag{B.22}$$

Weights w_i^{ent} emphasize lower-score nodes, and the stopped teacher prevents mutually changing views from amplifying dropout noise. The full objective is

$$\mathcal{L}_{\text{TTA}} = \lambda_{\text{ent}}\mathcal{L}_{\text{ent}} + \lambda_{\text{anc}}\mathcal{L}_{\text{anc}} + \lambda_{\text{cons}}\mathcal{L}_{\text{cons}}. \tag{B.23}$$

Algorithm 1 WaveGuard

Inputs: Unlabeled target graph $\mathcal{G}^t = (\mathbf{X}^t, \mathbf{A}^t)$; pre-trained source model f_{θ_0}; source domain spectral memory $\mathcal{M}$; set of updatable parameters Θ_{adapt}; number of adaptation steps T.

Output: Target node prediction $\widehat{\mathbf{Y}}^t$.

1. Initialize $\theta \leftarrow \theta_0$.
2. Compute frozen representations $\mathbf{H}^t$ and predictions $\mathbf{P}^0$ on the original target graph.
3. Compute spectral deviations D_i and fixed scores ρ_i using Equations (B.2)–(B.14).
4. Compute the directional edge weights $\omega_{j\to i}$ and construct the directed message-weight matrix $\mathbf{W}^t$.
5. Compute frozen predictions $\mathbf{P}^{\text{orig}}$, $\mathbf{P}^{\text{pur}}$ and anchor reliability r_i^{anc}.
6. For $t = 1, \ldots, T$:
   1. Perform a forward pass of the current model on $\mathbf{W}^t$ and generate two independent Dropout views;
   2. Compute weighted entropy, anchor, and consistency losses using Equations (B.21)–(B.22);
   3. Update $\theta \leftarrow \theta - \eta_{\text{opt}}\nabla_{\Theta_{\text{adapt}}}\mathcal{L}_{\text{TTA}}$.
7. Return $\widehat{\mathbf{Y}}^t = \text{argmax}_c f_\theta(\mathbf{X}^t, \mathbf{W}^t)_{:,c}$.

B.5 Computational Complexity and Storage

Computational complexity. For N_t target nodes, $|\mathcal{E}^t|$ support edges, hidden dimension d_h, K scales, order R, C classes, and B buckets, sparse wavelets cost $O(KR|\mathcal{E}^t|d_h)$, conditional scores $O(N_tCBK)$, and edge similarities $O(|\mathcal{E}^t|d_h)$. Sorting neighborhood similarities costs $O(\sum_i d_i \log d_i)$; directional weighting costs $O(|\mathcal{E}^t|)$.

Scores, purification, and anchors are computed once per target. The T updates add only constant-factor forward-pass overhead for two dropout views. Memory costs $O(CBK + B)$, independently of source-graph size, and excludes per-node records.

C Optimization Analysis

C.1 Stability of an Idealized Fixed-Score Surrogate Objective

Idealized analysis. We consider a fixed-score surrogate objective with fixed intervention quantities and a consistency teacher independent of the current parameters. Let ξ denote sampled views; define

$$\overline{\mathcal{L}}(\theta) = \mathbb{E}_\xi\left[\mathcal{L}_{\text{sur}}(\theta;\xi)\right]. \tag{C.1}$$

Here $\mathcal{L}_{\text{sur}}$ uses the loss terms in Equation (B.23) with fixed ρ_i, $\mathbf{W}^t$, $\mathbf{p}_i^{\text{pur}}$, r_i^{anc}, and consistency teachers. The following bound assumes unbiased SGD updates of this idealized objective.

Proposition C.1 (Stability of an Idealized Fixed-Score Surrogate Objective). Assume: (1) the expected objective $\overline{\mathcal{L}}(\theta)$ is L_θ-smooth with respect to the updatable parameters Θ_{adapt}, and there exists a finite lower bound $\overline{\mathcal{L}}_*$; (2) the stochastic gradient $\mathbf{g}_t$ at each step satisfies $\mathbb{E}[\mathbf{g}_t \mid \theta_t] = \nabla\overline{\mathcal{L}}(\theta_t)$ and $\mathbb{E}[\| \mathbf{g}_t - \nabla\overline{\mathcal{L}}(\theta_t) \|_2^2 | \theta_t] \le \sigma_g^2$; (3) the optimization step size satisfies $0 < \eta_{\text{opt}} \le 1/L_\theta$. After T SGD updates, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\| \nabla\overline{\mathcal{L}}(\theta_t) \|_2^2\right] \le \frac{2\left[\overline{\mathcal{L}}(\theta_0) - \overline{\mathcal{L}}_*\right]}{\eta_{\text{opt}}T} + L_\theta\eta_{\text{opt}}\sigma_g^2. \tag{C.2}$$

Equation (C.2) bounds the average expected squared gradient norm by a term of order $O(1/T)$ plus the noise term $L_\theta \eta_{\text{opt}} \sigma_g^2$. With a fixed step size and stochastic gradients, this bound permits a nonzero noise floor. With deterministic gradients, the noise term vanishes and the average squared gradient norm converges to zero. This is not a convergence guarantee for the actual Adam algorithm in Appendix D.2, whose consistency teacher depends on current predictions. It guarantees neither a global optimum nor target classification risk reduction; Section 5 reports empirical adaptation dynamics.

D Experimental and Reproducibility Details

D.1 Datasets, Domains, and Evaluation Protocol

Evaluation protocol. Each target starts independently from the same source checkpoint under the source-free, transductive protocol in Appendix A.1. Only target features and undirected structure are available during adaptation; target labels are reserved for final evaluation.

Datasets. CSBM (Deshpande et al., 2018) provides 8 scenarios varying class priors, connection strength, and homophily; each graph has 6,000 nodes, 3 classes, and 20-dimensional features. ogbn-arxiv (Hu et al., 2020) uses source cutoffs 2007/2009/2011 and target windows 2015–2016/2017–2018. Elliptic (Weber et al., 2019) uses temporal partitions E1–E3; time_step is excluded from the 165 input features, and "unknown" nodes are excluded from supervised loss and evaluation.

PubMed (Yang et al., 2016) constructs equal-sized source and target pools by class, replacing target-related edges at $r = 0.30/0.50/0.60$. Twitch (Rozemberczki et al., 2021) transfers from DE to ENGB, ES, and FR independently. Source nodes use class-stratified 60%/20%/20% train/validation/test splits. Table D1 gives partitions and core configurations.

Table D1: Datasets, domain settings, and core evaluation configurations.

Dataset	Source / training	Target / shift	Backbone	Updates / LR / steps	Metric
CSBM	Synthetic source graph; 60/20/20	CSBM1–CSBM8	GSN	classifier / 1e-3 / 100	Accuracy
ogbn-arxiv	Year $\leq$ 2007/ 2009/2011	2015–2016; 2017–2018	GCN	classifier / 1.2e-3 / 600	Accuracy
Elliptic	Time Steps 1–22	E1:29–34; E2:35–42;E3:43–49	GPRGNN	all / 1e-3 / 140	Balanced Accuracy; Macro-F1
PubMed	50% source pool; 60/20/20	3% adaptation / 97% evaluation; r=0.30/0.50/0.60	GPRGNN	BN / 5e-4 (150) or 3e-4 (200)	Accuracy
Twitch	DE; 60/20/20	ENGB; ES; FR	GPRGNN	all / 1e-3 / 100	Accuracy

The 8 CSBM scenarios, 6 ogbn-arxiv windows, 3 Elliptic periods, 3 PubMed perturbations, and 3 Twitch targets give 23 target scenarios.

D.2 Baselines, Implementation, Hyperparameters, and Seeds

Baselines. We compare TENT (Wang et al., 2021), T3A (Iwasawa & Matsuo, 2021), LAME (Boudiaf et al., 2022), ActMAD (Mirza et al., 2023), GTrans (Jin et al., 2023), SOGA (Mao et al., 2024), HomoTTT (Zhang et al., 2024a), T3R (Truong et al., 2026), TTReFT (Zhang et al., 2026), GraphPatcher (Ju et al., 2023), GOAT (Zhang et al., 2025), and Matcha (Bao et al., 2025)/TSA (Hsu et al., 2026) combined with T3A, TENT, or LAME. Preprocessing, checkpoints, target masks, and metrics are shared; original source-training and unsupervised stopping procedures are retained.

Implementation. CSBM uses GSN (GraphSAGE-style aggregation) (Hamilton et al., 2017), ogbn-arxiv uses GCN (Kipf & Welling, 2017), and other datasets use GPRGNN (Chien et al., 2021). Source ERM uses cross-entropy for 400 epochs with Adam (lr=0.003, weight decay=0),

dropout=0.5, and GPR order K=5. Test-time Adam optimizes the whole graph with the scopes, learning rates, and steps in Table D1.

Configuration. Defaults are B=10, ξ=0.05, $\omega_{\min}$=0.1, γ=2, $\lambda_{\text{ent}} = 1.0$, $\lambda_{\text{anc}} = 0.2$, and $\lambda_{\text{cons}} = 0.5$. Scores, graphs, and anchors are fixed per target. The frozen model stays in evaluation mode; final evaluation disables dropout. Hyperparameters use source validation or an independent development domain, never target evaluation labels.

Reproducibility. Main comparison means and standard deviations use five seeds {99,30,50,23,73} and the metrics in Table D1. Software versions are Python 3.10, PyTorch 2.2.2, PyTorch Geometric 2.5.2, and CUDA 12.1.

D.3 Additional CSBM, ogbn-arxiv, and Elliptic Results

Figure 3 reports approximately 79.5% mean accuracy across eight CSBM scenarios, with gains over ERM in every setting. Gains reach 33.5, 33.4, 22.0, and 20.2 percentage points on CSBM2, CSBM5, CSBM6, and CSBM4. On CSBM1 and CSBM3, 91.2% $\pm$ 1.93% and 88.6% $\pm$ 2.15% exceed the highest plotted baseline means by 1.2 and 0.4 points.

WaveGuard leads the plotted methods on CSBM1–3, CSBM5, CSBM7, and CSBM8. It ranks second on CSBM4 and CSBM6: 74.3% versus 82.5%, and 58.5% versus 78.7%, leaving gaps of 8.2 and 20.2 points. Gains over ERM therefore coexist with scenario-dependent rankings among adaptation methods.

On ogbn-arxiv, WaveGuard's 54.11% average exceeds SOGA by 0.75 points. It leads on Arxiv1–3 and Arxiv5, ranks second on Arxiv4, and obtains 56.47% on Arxiv6 against SOGA's 59.46%.

On Elliptic, WaveGuard leads both metrics in all three periods. Average Balanced Accuracy and Macro-F1 are 75.14% and 66.70%, exceeding Matcha-LAME by 6.12 and 6.63 points. Macro-F1 gains on Elliptic2 and Elliptic3 are 5.00 and 5.70 points, supporting improvements beyond majority-class prediction.

D.4 Additional PubMed and Twitch Results

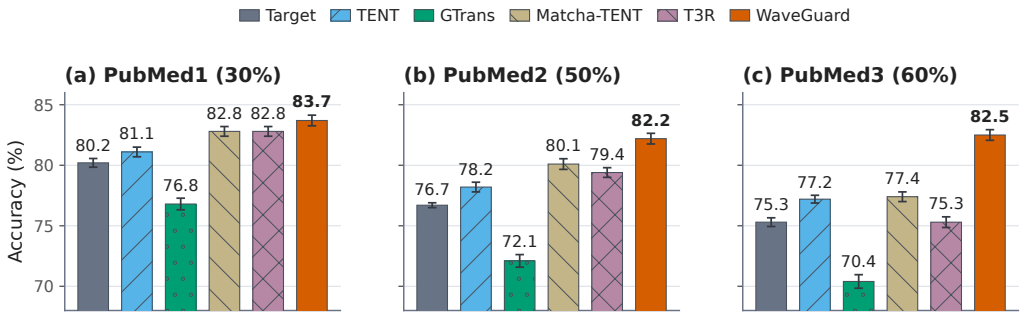

Figure D1: Target node classification accuracy on PubMed at edge replacement rates of 30%, 50%, and 60% (%, mean $\pm$ standard deviation). The plotted methods are Target, TENT, GTrans, Matcha-TENT, T3R, and WaveGuard.

As PubMed edge replacement rises from 30% to 60%, source accuracy falls from 80.2% to 75.3%. WaveGuard obtains 83.7%, 82.2%, and 82.5%, with gains of 3.5, 5.5, and 7.2 points. Its 82.8% average exceeds Matcha-TENT and T3R by 2.7 and 3.6 points; it leads at all three intensities. The more precise endpoint gains are 3.56 and 7.19 points.

Table 3 gives all Twitch comparisons. WaveGuard leads on DE→ENGB (57.46%) and DE→FR (56.88%), improving over the source model by 1.54 and 16.53 points. Its 56.34% average is 0.11 points below TSA_T3A.

On DE→ES, the gain over the source model is 19.83 points, but WaveGuard trails ActMAD and Matcha-LAME. This target-dependent result does not identify the dominant shift type; that requires controlled decomposition or explicit distribution statistics.

Table D2: Balanced Accuracy and Macro-F1 for the three target time periods in Elliptic (%, mean ± standard deviation). Cell format: "Balanced Accuracy / Macro-F1."

Method	Elliptic1	Elliptic2	Elliptic3	Avg
ERM (Target)	79.68 ± 0.42 / 72.30 ± 0.50	43.84 ± 1.15 / 32.10 ± 1.10	42.63 ± 0.88 / 30.40 ± 0.90	55.38 / 44.93
TENT	83.24 ± 0.55 / 76.80 ± 0.58	47.71 ± 1.32 / 36.20 ± 1.30	46.47 ± 0.96 / 34.50 ± 1.00	59.14 / 49.17
TSA_TENT	83.24±0.48 / 76.70±0.55	47.71±1.24 / 36.10±1.25	46.43 ± 1.05 / 34.40 ± 1.05	59.13 / 49.07
TTReFT	81.66 ± 0.63 / 74.90 ± 0.65	41.29 ± 1.41 / 29.20 ± 1.40	38.60 ± 1.12 / 26.00 ± 1.15	53.85 / 43.37
GTrans	79.64 ± 0.37 / 73.10 ± 0.45	45.12 ± 0.95 / 34.20 ± 0.95	43.51 ± 0.74 / 31.80 ± 0.80	56.09 / 46.37
LAME	79.64 ± 0.51 / 73.00 ± 0.60	40.63 ± 1.28 / 28.80 ± 1.30	41.24 ± 1.19 / 29.50 ± 1.20	53.84 / 43.77
TSA_LAME	78.94 ± 0.44 / 72.30 ± 0.52	41.32 ± 1.36 / 29.50 ± 1.40	41.88 ± 1.08 / 30.20 ± 1.10	54.05 / 44.00
HomoTTT	77.93 ± 0.67 / 71.20 ± 0.75	40.49 ± 1.45 / 28.00 ± 1.50	38.60 ± 1.23 / 25.80 ± 1.30	52.34 / 41.67
SOGA	76.39 ± 0.72 / 69.60 ± 0.80	39.79 ± 1.38 / 27.50 ± 1.40	38.20 ± 1.31 / 25.40 ± 1.35	51.46 / 40.83
ActMAD	75.16 ± 0.85 / 68.30 ± 0.95	38.43 ± 1.52 / 26.00 ± 1.60	36.37 ± 1.42 / 23.50 ± 1.50	49.99 / 39.27
T3A	64.94 ± 1.24 / 57.20 ± 1.35	26.48 ± 1.86 / 16.20 ± 1.90	16.61 ± 2.15 / 7.50 ± 2.20	36.01 / 26.97
Matcha-TENT	82.97±0.49 / 76.30±0.55	41.48 ± 1.17 / 29.80 ± 1.20	40.40 ± 0.92 / 28.20 ± 1.00	54.95 / 44.77
Matcha-LAME	78.50 ± 0.58 / 71.60 ± 0.65	56.46 ± 0.83 / 45.80 ± 0.90	72.10 ± 0.65 / 62.80 ± 0.75	69.02 / 60.07
Matcha-T3A	76.26±0.76 / 69.40±0.85	36.32 ± 1.64 / 24.20 ± 1.70	42.71 ± 1.03 / 30.50 ± 1.10	51.76 / 41.37
TSA_T3A	64.77 ± 1.31 / 57.00 ± 1.40	26.55 ± 1.79 / 16.30 ± 1.85	16.65 ± 2.08 / 7.60 ± 2.15	35.99 / 26.97
WaveGuard (Ours)	86.88 ± 0.39 / 80.80 ± 0.46	60.99 ± 0.76 / 50.80 ± 0.85	77.56 ± 0.58 / 68.50 ± 0.65	75.14 / 66.70

D.5 Score-Guided Selectivity and Long-Horizon Adaptation Details

D.5.1 Score-Guided Selective Edge Suppression

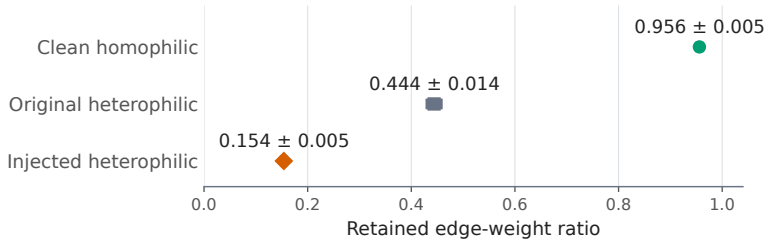

Figure D2: Retained edge weights on PubMed3 (60% edge replacement). Points show the cross-run mean of the within-run median retained-weight ratio, and error bars show one standard deviation over five runs. Clean homophilic edges retain the largest weights, while injected heterophilic edges receive the strongest attenuation.

D.5.2 Long-Horizon Adaptation Stability

The 30-step CSBM8 stress test compares Full WaveGuard, objective calibration alone, purification fixed after Step 1, entropy minimization, and the source model.

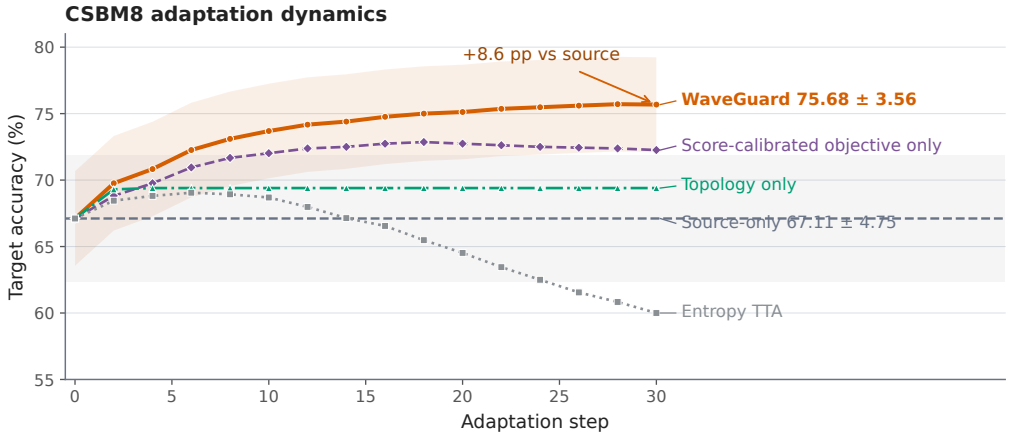

Figure D3: 30-step target accuracy trajectories for different adaptation strategies on CSBM8. The curves represent the average of five runs, with the shaded area indicating one standard deviation; the dashed line represents the unadapted baseline model, while Full WaveGuard employs both soft topology purification and the score-calibrated objective.

Full WaveGuard reaches 75.68% $\pm$ 3.56%, exceeding the source model's 67.11% $\pm$ 4.75% by 8.57 points. Both isolated interventions improve on the source model but trail the full method. Entropy minimization improves briefly, then degrades below the source model.

Purification attenuates higher-score messages, while calibration downweights their entropy and consistency terms through $(1-\rho_i)^\gamma$. Anchor weights instead use $\rho_i^\gamma r_i^{\mathrm{anc}}$. The shared score couples these complementary effects, consistent with the ablations in Section 5.4.

D.6 Extended Ablation Results

Figure 5 and Table D3 report component ablations on CSBM1, CSBM2, CSBM5, PubMed3, Elliptic1, Elliptic3, and Twitch1. For each scenario, the drop is the performance of full WaveGuard minus that of the ablation variant, expressed in percentage points. The cross-scenario average is the unweighted mean of the seven scenario-level mean drops.

Removing score guidance produces the largest mean drop in every scenario, ranging from 3.52 points on CSBM1 to 12.26 on Elliptic3. The drops are 10.06 and 9.80 on CSBM2 and CSBM5, 8.12 on PubMed3, 9.20 on Elliptic1, and 4.20 on Twitch1. This pattern supports the role of the shared intervention score across synthetic, temporal, and cross-lingual shifts.

The two intervention branches also contribute separately. Removing the score-calibrated objective and purification costs 3.67 and 2.87 points on average, respectively. Degree conditioning has its largest effects on Elliptic1 and Elliptic3, with drops of 5.00 and 7.36 points, and a seven-scenario average of 3.61. Replacing multiscale wavelets with a single scale and removing the anchor loss reduces performance by 1.98 and 2.17 points on average. All six variants reduce mean performance on Twitch1 as well, extending the component analysis to cross-lingual transfer. Hyperparameter sensitivity remains evaluated separately on CSBM2, PubMed3, and Elliptic3 in Section D.7.

Table D3: Complete component-ablation results. Entries are performance drops in percentage points (mean ± standard deviation) relative to full WaveGuard. Bold values identify the largest mean drop within each scenario. The last row gives the unweighted average of the seven mean drops.

Scenario	Single-scale	No degree cond.	No purification	No score-cal. TTA	No score guidance	No anchor loss
CSBM1	1.52 ± 0.21	1.12 ± 0.03	0.92 ± 0.03	1.32 ± 0.07	$\mathbf{3.52 \pm 0.29}$	0.52 ± 0.01
CSBM2	1.76 ± 0.14	3.66 ± 0.07	3.16 ± 0.11	4.46 ± 0.16	$\mathbf{10.06 \pm 1.31}$	2.36 ± 0.24
CSBM5	1.90 ± 0.05	3.80 ± 0.08	3.40 ± 0.11	4.20 ± 0.18	$\mathbf{9.80 \pm 1.15}$	2.40 ± 0.13
PubMed3	1.52 ± 0.19	2.82 ± 0.13	1.72 ± 0.09	3.22 ± 0.13	$\mathbf{8.12 \pm 0.71}$	1.42 ± 0.08
Elliptic1	2.70 ± 0.22	5.00 ± 0.17	3.80 ± 0.07	4.20 ± 0.11	$\mathbf{9.20 \pm 0.84}$	3.10 ± 0.15
Elliptic3	3.46 ± 0.17	7.36 ± 0.24	5.06 ± 0.21	5.76 ± 0.30	$\mathbf{12.26 \pm 2.11}$	3.96 ± 0.12
Twitch1	1.00 ± 0.34	1.50 ± 0.06	2.00 ± 0.11	2.50 ± 0.21	$\mathbf{4.20 \pm 0.51}$	1.40 ± 0.06
Average	1.98	3.61	2.87	3.67	8.17	2.17

D.7 Hyperparameter Sensitivity

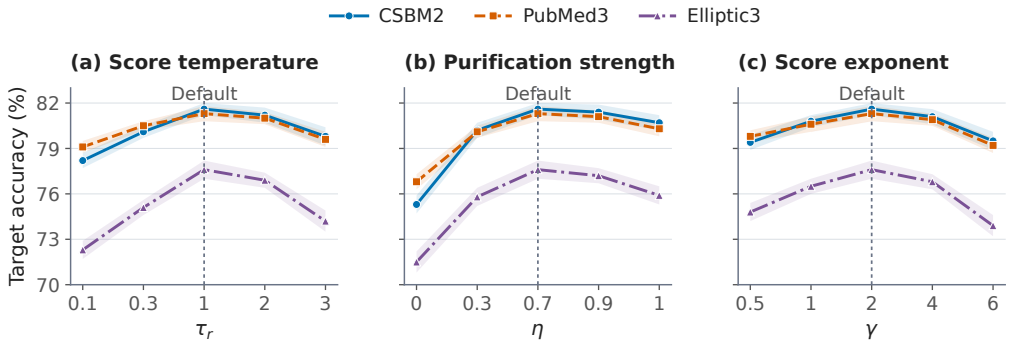

Figure D4: Sensitivity of WaveGuard to intervention-score temperature τ_r, soft topology purification strength η, and score exponent γ. The curves represent the mean of five runs, the shading indicates one standard deviation, and the vertical dashed line denotes the default configuration; the test scenarios are CSBM2, PubMed3, and Elliptic3.

Small τ_r makes gating nearly hard; large τ_r obscures score differences. Weak purification η barely changes weights; excessive η suppresses normal information. Small γ weakly differentiates updates; large γ suppresses useful signals.

Across all three scenarios, stable ranges are $\tau_r = 1$–2, $\eta = 0.7$–0.9, and $\gamma = 2$–4. Defaults $(\tau_r, \eta, \gamma) = (1.0, 0.7, 2)$ lie within these ranges and use source validation or an independent development domain. Figure D4 does not systematically vary τ_a, so no sensitivity conclusion is drawn for it.

E Additional Related-Work Comparison

E.1 Capability Matrix for Representative Adaptation Approaches

Table E1 compares information access, control granularity, and intervention points. "Partial" denotes capabilities available only in some methods of a class. WaveGuard shares one node-wise score between message weighting and objective calibration.

Table E1: Comparison of adaptation approaches. Src.: source access during adaptation; Score: node-wise intervention; Spec.: multiscale spectrum; Cond.: class/degree conditioning; Topo.: topology intervention; Upd.: parameter updates; Dir.: directional gating.

Approach / Method	**Src.**	**Score**	**Spec.**	**Cond.**	**Topo.**	**Upd.**	**Dir.**
Graph domain adaptation (Liu et al., 2023; 2024a;b; You et al., 2023; Cai et al., 2024)	R	N	P	N	P	Y	N
Graph OOD generalization (Sui et al., 2023; Li et al., 2024a;b; Wu et al., 2024; Guo et al., 2024; Yu et al., 2023; Ma, 2024a)	–	N	P	N	Tr	N	N
Generic TTA (Mirza et al., 2023; Wang et al., 2021; Iwasawa & Matsuo, 2021; Boudiaf et al., 2022)	N	P	N	N	N	M	N
Model-centric GTTA (Mao et al., 2024; Zhao et al., 2025; Zhang et al., 2024a; Truong et al., 2026; Zhang et al., 2026)	N	P	N	N	N	Y	N
Data-centric GTTA (Jin et al., 2023; Zhang et al., 2024b; 2025; Luo et al., 2024a;b)	N	P	N	N	Y	P	N
GraphPatcher (Ju et al., 2023)	N	N	N	D	Aug	N	N
Matcha (Bao et al., 2025)	N	N	N	N	Agg	TTA	N
TSA (Hsu et al., 2026)	N	P	N	N	NW	TTA	N
WaveGuard	**S**	**Y**	**Y**	**C+D**	**Soft**	**Y**	**Y**

Y/N: yes/no; P: partial; R: required; S: compact summaries only; –: no adaptation phase; C/D: class/degree. Topology codes: Tr, during training; Aug, graph augmentation; Agg, aggregation adjustment; NW, neighborhood weighting; Soft, soft topology purification. Update codes: M, method-dependent; TTA, combined with TTA.

F Enlarged Views of the Control Mechanisms

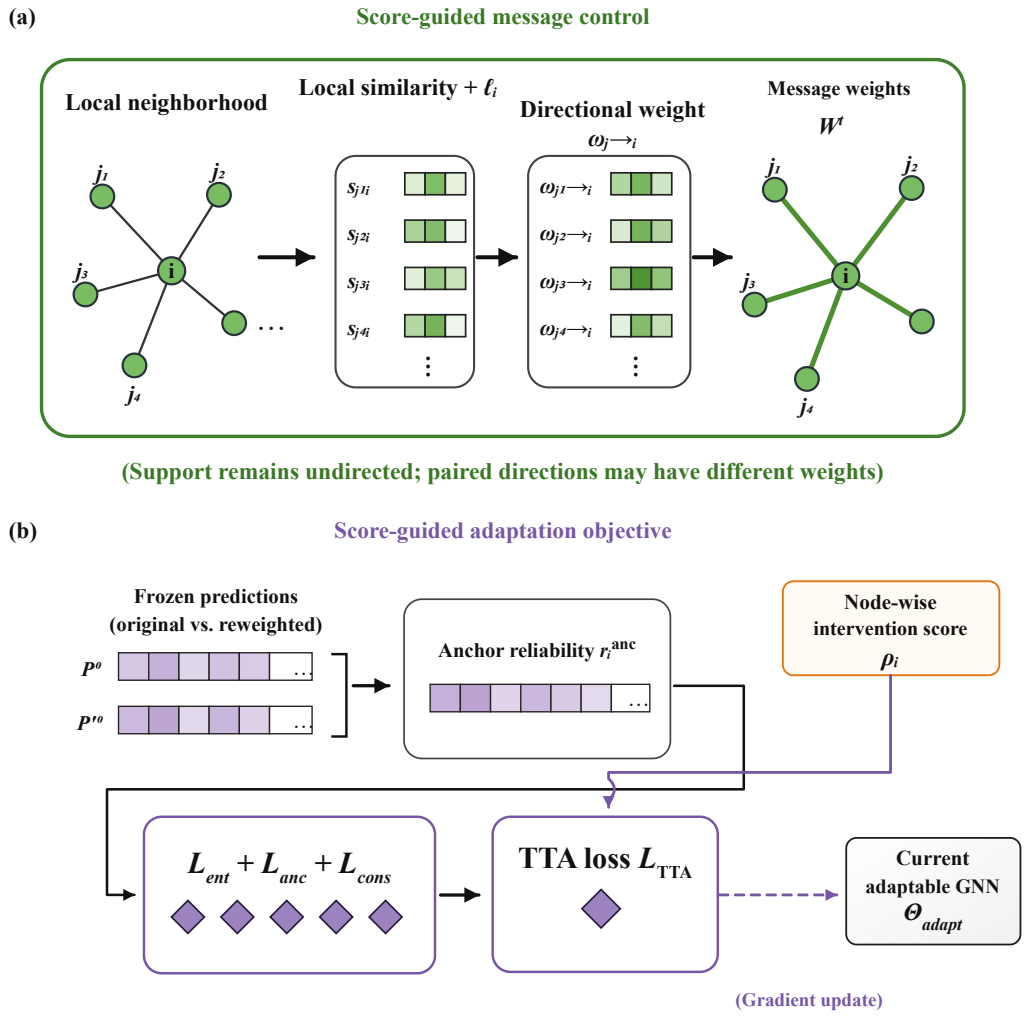

Figure F1: Enlarged views of the control branches in Figure 2. (a) Local similarities and the receiver-specific threshold ℓ_i determine directional message weights together with the fixed intervention scores (Appendix B.3). (b) Frozen predictions determine anchor reliability, and the fixed score calibrates the adaptation objective (Appendix B.4).

Notation. In Figures 2 and F1, $\mathbf{P}^0$ and $\mathbf{P}'^0$ denote the frozen prediction matrices on the original and reweighted target graphs, respectively; their rows are $\mathbf{p}_i^{\text{orig}}$ and $\mathbf{p}_i^{\text{pur}}$ in Equation (B.18). The displayed $\mathbf{W}^t$ denotes the directed message-weight matrix in Equation (B.17). The similarity entries $s_{j_k i}$ abbreviate $s_{j_k\to i}$ in Equation (B.15).

Source memory and target scoring. In Figure 2, controlled edge dropout constructs source-side scale weights (Appendix B.1). The target-path fusion symbol combines spectral deviation and entropy through the thresholded sigmoid in Equation (B.14).

Objective weights. The displayed loss sum abbreviates Equation (B.23), including its coefficients $\lambda_{\text{ent}}, \lambda_{\text{anc}}, \lambda_{\text{cons}}$. Entropy and consistency terms use $(1-\rho_i)^\gamma$, whereas the anchor term uses $\rho_i^\gamma r_i^{\text{anc}}$. Scores, reweighted graphs, and frozen anchors remain fixed; dashed purple arrows denote updates to Θ_{adapt}.

G COMPLETE OGBN-ARXIV RESULTS

Table G1 provides the means and standard deviations for the comparison summarized in Table 1. The two panels group the six target windows for readability.

Table G1: Target node classification accuracy on ogbn-arxiv (%, mean ± standard deviation). Panel (a) reports Arxiv1–Arxiv3; panel (b) reports Arxiv4–Arxiv6 and the six-window average. The best result is bolded and the second-best is underlined.

(a)

Method	**Arxiv1**	**Arxiv2**	**Arxiv3**
ERM (Target)	45.77±0.42	48.37±0.51	53.86±0.47
ActMAD	45.73±0.45	48.32±0.49	53.95±0.53
GTrans	45.71±0.36	47.92±0.41	53.99±0.44
HomoTTT	42.71±0.98	43.67±1.12	49.14±1.05
SOGA	42.89±2.86	49.47±3.15	53.41±2.94
LAME	44.72±0.82	49.87±0.91	54.11±0.87
T3A	43.35±1.02	45.60±1.14	51.86±1.07
TENT	45.24±0.33	47.72±0.38	53.83±0.35
TSA_LAME	<u>45.94</u>±0.86	<u>50.28</u>±0.95	54.34±0.90
TSA_T3A	44.75±1.05	47.64±1.17	53.35±1.10
TSA_TENT	45.58±0.37	47.89±0.43	<u>54.72</u>±0.39
TTReFT	34.06±3.24	26.91±3.61	37.30±3.35
T3R	33.36±3.47	27.81±3.83	32.50±3.56
WaveGuard	**46.36**±0.41	**51.87**±0.46	**55.07**±0.43

(b)

Method	**Arxiv4**	**Arxiv5**	**Arxiv6**	**Avg.**
ERM (Target)	53.23±0.55	58.37±0.62	55.73±0.58	52.5550±0.48
ActMAD	53.69±0.51	58.76±0.57	56.11±0.54	52.7600±0.50
GTrans	53.62±0.42	58.53±0.48	56.45±0.46	52.7033±0.42
HomoTTT	47.43±1.20	56.39±1.14	53.66±1.08	48.8333±1.06
SOGA	**57.20**±3.22	57.68±3.07	**59.46**±3.31	<u>53.3517</u>±3.02
LAME	53.98±0.94	58.82±0.90	54.97±0.85	52.7450±0.88
T3A	50.08±1.18	55.25±1.11	52.04±1.04	49.6967±1.09
TENT	53.90±0.40	58.51±0.37	56.97±0.39	52.6950±0.36
TSA_LAME	54.86±0.97	58.73±0.93	54.97±0.89	53.1867±0.91
TSA_T3A	53.08±1.21	56.22±1.15	53.79±1.08	51.4717±1.12
TSA_TENT	54.55±0.45	<u>58.97</u>±0.41	<u>56.98</u>±0.44	53.1150±0.40
TTReFT	30.50±3.72	24.19±3.48	13.15±3.86	27.6850±3.51
T3R	19.13±3.95	44.65±3.71	37.16±3.64	32.4350±3.68
WaveGuard	<u>55.27</u>±0.48	**59.59**±0.44	56.47±0.47	**54.1050**±0.44