JOINT DOMAIN ADAPTATION AND UNKNOWN CLASS RECOGNITION ON HETEROPHILIC GRAPHS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph domain adaptation transfers knowledge from a labeled source graph to an unlabeled target graph. This task becomes challenging when graphs exhibit heterophily and the target contains classes absent from the source. Existing methods largely address these conditions separately, leaving the joint problem insufficiently explored. We propose HODA, a unified framework for **H**eterophilic **O**pen-set **D**omain **A**daptation. HODA uses multiscale spectral encoding to preserve complementary low-pass and high-pass information across propagation orders. Using the resulting representations, we adapt source class prototypes to the target distribution and introduce an unknown prototype for target private classes. Agreement between classifier predictions and prototype-based clustering selects target pseudo labels and excludes nodes jointly predicted as unknown from domain alignment. We also identify and correct inconsistent label encodings across the processed WebKB graphs, enabling semantically consistent cross domain evaluation. Experiments across multiple datasets and transfer directions demonstrate competitive performance in known-class classification and unknown-class recognition.

1 INTRODUCTION

Graphs represent entities and their relationships in social, citation, and biological networks (Hu et al., 2020; Dwivedi et al., 2023). Node classification is a fundamental task in these systems, but obtaining labels for every new graph is costly (Hamilton et al., 2017). Graph domain adaptation (GDA) addresses this challenge by transferring knowledge from a labeled source graph to an unlabeled target graph (Wu et al., 2020; Cai et al., 2024). However, transfer becomes challenging under heterophily, where connected nodes may belong to different classes (Zhu et al., 2020; Luan et al., 2022). Neighborhood patterns may also vary across domains, limiting the transferability of structural information (Liu et al., 2024; Fang et al., 2025). Moreover, target graphs may contain classes absent from the source, requiring unknown-class recognition alongside known-class classification (Wang et al., 2024; Shen et al., 2025; Zhang & He, 2026). We study this joint problem of open-set GDA under heterophily.

Existing work addresses these challenges largely separately. Open-set GDA methods use uncertainty-based grouping, selective alignment, or graph and model reprogramming to distinguish known and unknown target nodes (Wang et al., 2024; Shen et al., 2025; Zhang & He, 2026), but do not explicitly address representation learning under heterophily. Meanwhile, heterophilic graph learning explores complementary neighborhood and spectral information (Zhu et al., 2020; Luan et al., 2022; Zhang et al., 2026). Recent GDA methods further address homophily differences through mixed filters or separately reconstructed graph views, but assume a shared label space across domains (Fang et al., 2025; 2026). Two challenges therefore remain: learning transferable representations from diverse neighborhood patterns, and obtaining reliable target supervision without forcing unknown nodes into known classes. The latter is particularly difficult because an uncertain prediction may indicate either a poorly transferred known-class or an unseen class.

To address these challenges, we propose HODA, a unified framework for open-set GDA under heterophily. We first use multiscale spectral encoding to preserve complementary information through low-pass and high-pass filtering across propagation orders. We then adapt deep embedded clustering (Xie et al., 2016) to transfer source class prototypes to the target domain. Classifier predictions and source-prototype assignments guide prototype refinement, while a dedicated unknown prototype

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.

represents potentially unseen target classes. Finally, classifier-clustering consensus selects pseudo labels for target self-training and excludes jointly predicted unknown nodes from domain alignment. We also identify inconsistent label encodings across the processed WebKB graphs released with Geom-GCN (Pei et al., 2020). Using the original semantic annotations (Craven et al., 1998), we establish a shared encoding for cross domain evaluation.

Our main contributions are:

- We formulate heterophilic open-set graph domain adaptation, where structural heterophily, domain shift, and target-private classes jointly create representation and negative-transfer challenges.
- We introduce HODA, which couples multiscale low/high-pass encoding with reliability-aware prototype adaptation and classifier-prototype consensus to identify reliable known and unknown target nodes before alignment.
- We provide theoretical and empirical analyses connecting multiscale representations, unknown contamination, consensus accuracy, and selective domain alignment.
- We identify and correct inconsistent semantic label mappings in commonly used WebKB cross-domain benchmarks and conduct extensive evaluation across heterophilic and homophilic graphs.

2 Related Work

2.1 Learning on Heterophilic Graphs

Early work improves neighborhood aggregation by separating ego and neighbor representations and incorporating higher-order neighborhoods (Zhu et al., 2020). Subsequent work explores complementary graph signals by adaptively combining aggregation, diversification, and identity channels to accommodate local graph patterns (Luan et al., 2022). More recently, HeroFilter extends this direction with adaptive spectral filtering, motivated by the finding that heterophily does not have a strictly monotonic relationship with spectral responses (Zhang et al., 2026). GRAPHITE instead introduces feature nodes to facilitate homophilic message passing between nodes with similar features (Qiu et al., 2026). These methods improve representation learning within a graph, but do not directly address distribution shifts between source and target graphs.

2.2 Graph Domain Adaptation on Heterophilic Graphs

Transferring knowledge across graphs requires accounting for differences in how nodes of different classes connect. PairAlign mitigates shifts in class-dependent connection patterns by reweighting edges to adjust neighbor contributions during aggregation (Liu et al., 2024). More recent work shows that differences in graph homophily can impair transfer and introduces mixed filters to mitigate these discrepancies (Fang et al., 2025). To capture both homophilic and heterophilic information, RSGDA reconstructs corresponding graph views, extracts complementary signals through adaptive filtering, and aligns these views across domains (Fang et al., 2026). Although these methods address structural differences across graphs, they assume a shared label space between domains and lack explicit mechanisms for recognizing target private classes.

2.3 Open-Set Graph Domain Adaptation

Open-set GDA allows the target graph to contain classes absent from the source graph, requiring both known-class classification and unknown-class recognition. SDA uses prediction uncertainty to partition target nodes and applies separate alignment objectives to different groups (Wang et al., 2024). UAGA further reduces negative transfer by aligning known target nodes with the source while pushing unknown nodes away (Shen et al., 2025). More recently, GraphRTA modifies target graph structure and node features to improve known-unknown separation and reprograms the model to reduce source-specific bias (Zhang & He, 2026). However, these methods either rely on homophily-based neighborhood aggregation or lack explicit designs for learning representations under heterophily.

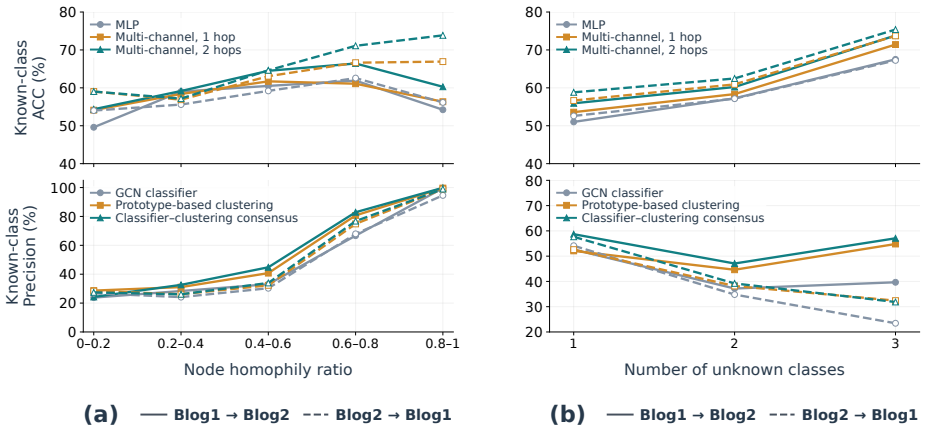


Figure 1: Performance of different encoders and prediction strategies on Blog. (a) Results across node homophily ranges with two unknown classes. (b) Results with varying numbers of unknown classes. The top row reports known-class accuracy for different encoders. The bottom row reports known-class precision for a GCN classifier, prototype-based clustering, and their prediction agreement. All results are averaged over five random seeds.

3 Motivation

3.1 Preliminaries

Let $\mathcal{G}_s = (\mathbf{A}_s, \mathbf{X}_s, \mathbf{y}_s)$ denote a labeled source graph with n_s nodes, where $\mathbf{A}_s \in \mathbb{R}^{n_s \times n_s}$ is the adjacency matrix, $\mathbf{X}_s \in \mathbb{R}^{n_s \times d}$ is the node feature matrix, and $\mathbf{y}_s = (y_{s,i})_{i=1}^{n_s}$ is the label vector with $y_{s,i} \in \mathcal{Y}_s = \{0, \dots, C-1\}$. The unlabeled target graph is denoted by $\mathcal{G}_t = (\mathbf{A}_t, \mathbf{X}_t)$, where $\mathbf{A}_t \in \mathbb{R}^{n_t \times n_t}$ and $\mathbf{X}_t \in \mathbb{R}^{n_t \times d}$. For each domain $a \in \{s, t\}$, let $\tilde{\mathbf{A}}_a = \mathbf{D}_a^{-1/2} \mathbf{A}_a \mathbf{D}_a^{-1/2}$, where $\mathbf{D}_a$ is the degree matrix. The low-pass and high-pass operators are $\mathbf{P}_{a,L} = (\mathbf{I}_{n_a} + \tilde{\mathbf{A}}_a)/2$ and $\mathbf{P}_{a,H} = (\mathbf{I}_{n_a} - \tilde{\mathbf{A}}_a)/2$, where $\mathbf{I}_{n_a}$ is the identity matrix. In the open-set setting, the target label space $\mathcal{Y}_t$ strictly contains the source label space. The target private classes form the set $\mathcal{Y}_u = \mathcal{Y}_t \setminus \mathcal{Y}_s$. For prediction and evaluation, all target private classes are grouped into a single unknown class indexed by C. Following Pei et al. (2020), we define the local node homophily ratio for a non-isolated node i as

$$h_i^{\text{node}} = \frac{|\{j \in \mathcal{N}(i) : y_j = y_i\}|}{|\mathcal{N}(i)|},$$

where $\mathcal{N}(i)$ is the neighbor set of node i. A lower node homophily ratio indicates that a larger proportion of the node's neighbors belong to different classes. Given the labeled source graph and the unlabeled target graph, our goal is to correctly classify target nodes from known classes while identifying nodes from target private classes as unknown. The predictor therefore maps each target node to the output label space $\mathcal{Y}_s \cup \{C\}$.

3.2 Empirical Insights

Complementary graph filters and propagation orders support heterophilic graph learning (Luan et al., 2022; Zhu et al., 2020), but their role in open-set adaptation remains unclear. We examine representation learning and prediction reliability on Blog across node homophily ranges and different numbers of unknown classes (Figure 1). We first compare an MLP with one-hop and two-hop multi-channel encoders. The benefit of propagation varies across homophily groups, while the two-hop encoder achieves the highest overall known-class accuracy among these encoders in both transfer directions under all three open-set settings. This motivates retaining complementary graph signals across multiple propagation orders. We then compare a source-trained classifier, prototype-based clustering, and their consensus, which retains nodes only when both predictors agree on a known class. Consensus achieves higher overall known-class precision than the classifier alone in both

transfer directions across the evaluated open-set settings. This supports using classifier-clustering agreement to select reliable target pseudo labels.

3.3 Theoretical Analysis

To provide theoretical support for these observations, we derive an open-set risk bound. For a fixed encoder, let $Z^{(J)} = (Z_I, Z_{L,1}, Z_{H,1}, \ldots, Z_{L,J}, Z_{H,J})$ denote the joint representation containing the identity, low-pass, and high-pass views up to propagation order J. Let $\mathcal{H}_J$ be the corresponding class of predictors mapping $Z^{(J)}$ to $\{0, \ldots, C\}$. We assume that additional views can be ignored, so that $\mathcal{H}_j \subseteq \mathcal{H}_J$ for $j \leq J$, with lower-order predictors identified through projection onto their required views. We use the loss $\ell(h(Z^{(J)}), Y) = \mathbb{I}(h(Z^{(J)}) \neq Y)$, where all target private labels are mapped to C. Expectations and probabilities indexed by s and t refer to the source and target domains.

Definition 1 (Open-set risk). *Let $\pi_u = \Pr_t(Y = C) \in (0, 1)$ denote the target unknown-class prior. For $h \in \mathcal{H}_J$, define $R_s(h) = \mathbb{E}_s[\ell(h(Z^{(J)}), Y)]$ and $R_t^{\mathrm{os}}(h) = \mathbb{E}_t[\ell(h(Z^{(J)}), Y)]$ as the source risk and target open-set risk. For each propagation order j, define the ideal joint known-class risk as $\lambda_j = \inf_{h' \in \mathcal{H}_j}\{\mathbb{E}_s[\ell(h'(Z^{(j)}), Y)] + \mathbb{E}_t[\ell(h'(Z^{(j)}), Y) \mid Y < C]\}$.*

Definition 2 (Consensus selection). *For fixed consensus selectors, let S_k and S_u denote the events that the classifier and prototype predictor agree on a known class and the unknown class. For S_k, let $\widetilde{Y}$ be the agreed label. Their coverage rates are $\rho_k = \Pr_t(S_k \mid Y < C)$ and $\rho_u = \Pr_t(S_u \mid Y = C)$. The known-class consensus error is $\epsilon_{\mathrm{con}} = \Pr_t(\widetilde{Y} \neq Y \mid S_k, Y < C)$, set to zero when $\Pr_t(S_k, Y < C) = 0$. For an evaluated predictor h, define $L_k(h) = \mathbb{E}_t[\mathbb{I}(S_k)\ell(h(Z^{(J)}), \widetilde{Y})]$ and $L_u(h) = \mathbb{E}_t[\mathbb{I}(S_u)\ell(h(Z^{(J)}), C)]$.*

Definition 3 (Selective alignment). *Let $P_s^{(J)}$ denote the source distribution of $Z^{(J)}$ and $Q_t^{(J)} = P_t(Z^{(J)} \mid S_u^c)$ the target distribution retained for alignment. Define $\mathcal{D}_J = \mathrm{MMD}_{k_J}(P_s^{(J)}, Q_t^{(J)})$, where k_J is a positive semidefinite kernel on the joint representation space satisfying $k_J(z, z) \leq \kappa_J^2$ for all z, with $\kappa_J > 0$. Let $\eta = \Pr_t(Y = C \mid S_u^c)$ denote residual unknown contamination and $\beta = \Pr_t(S_u \mid Y < C)$ the known-class exclusion rate.*

Theorem 1 (Multiscale open-set risk bound). *Assume $\Pr_t(S_u^c) > 0$ and $\beta < 1$. For every $h, h' \in \mathcal{H}_J$, suppose that the disagreement function $d_{h,h'}(z) = \mathbb{I}(h(z) \neq h'(z))$ belongs to the RKHS of k_J with norm at most $B_J > 0$. Then, for every $h \in \mathcal{H}_J$,*

$$R_t^{\mathrm{os}}(h) \leq \min\Big\{(1 - \pi_u)\big[R_s(h) + B_J\mathcal{D}_J + 2B_J\kappa_J(\eta + \beta) + \lambda_J\big], \\ L_k(h) + (1 - \pi_u)\big[\rho_k\epsilon_{\mathrm{con}} + 1 - \rho_k\big]\Big\} + L_u(h) + \pi_u(1 - \rho_u). \tag{1}$$

Moreover, under the nested predictor-class assumption, $\lambda_J \leq \lambda_j$ for every $j \leq J$.

The proof is provided in Appendix A.6.1. The relation $\lambda_J \leq \lambda_j$ shows that retaining additional scales cannot increase the ideal joint risk when unhelpful views can be ignored. Meanwhile, $\mathcal{D}_J$ measures cross domain discrepancy in the joint multiscale representation. The terms η and β highlight the need to limit unknown contamination without incorrectly excluding known nodes from alignment. The terms ϵ_{con} and $L_k(h)$ suggest that consensus supervision should provide accurate pseudo labels and support effective fitting. Similarly, $L_u(h)$ highlights the need to learn from selected unknown nodes while identifying a sufficient proportion of the unknown population. These observations motivate multiscale encoding, consensus-guided supervision, and selective alignment.

4 Methodology

4.1 Multiscale Spectral Encoding

Figure 2 provides an overview of HODA. For each domain $a \in \{s, t\}$, we construct multiscale views $\mathbf{X}_{a,I} = \mathbf{X}_a$, $\mathbf{X}_{a,L,j} = \mathbf{P}_{a,L}^j\mathbf{X}_a$, and $\mathbf{X}_{a,H,j} = \mathbf{P}_{a,H}^j\mathbf{X}_a$ for $j = 1, \ldots, J$. Low-pass views capture neighborhood similarities, while high-pass views emphasize differences between connected nodes.

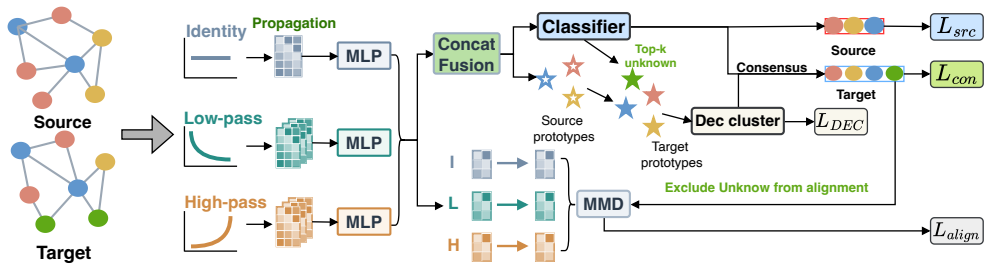


Figure 2: Overview of HODA. Multiscale spectral views are independently encoded for classification and prototype-based clustering. Source prototypes are adapted using target predictions and source-prototype assignments. Classifier-clustering agreement selects target pseudo labels, while consensus unknown nodes are excluded from channel-wise domain alignment.

Retaining different propagation orders provides information from multiple neighborhood ranges. Each view is processed by a separate MLP, with parameters shared across domains:

$$\mathbf{H}_{a,I} = \mathrm{MLP}_I(\mathbf{X}_a), \quad \mathbf{H}_{a,r,j} = \mathrm{MLP}_{r,j}(\mathbf{X}_{a,r,j}), r \in \{L, H\}. \tag{2}$$

We concatenate the encoded views and obtain the node representations through a shared fusion mapping:

$$\mathbf{Z}_a = F\left(\mathbf{H}_{a,I} \| \mathbf{H}_{a,L,1} \| \mathbf{H}_{a,H,1} \| \cdots \| \mathbf{H}_{a,L,J} \| \mathbf{H}_{a,H,J}, \mathbf{X}_a\right), \tag{3}$$

where $\|$ denotes concatenation and F is a shared fusion mapping with an attribute residual and layer normalization. Let $\mathbf{z}_{a,i}$ denote the fused representation of node i in $\mathbf{Z}_a$. A shared $(C+1)$-class softmax classifier produces $\mathbf{P}_a$, whose i-th row $\mathbf{p}_{a,i} = (p_{a,i,0}, \ldots, p_{a,i,C})$ contains the class probabilities for node i.

To explain the role of multi-hop high-pass filtering under heterophily, we analyze its effect on class information and spectral interference in a simplified graph model.

Theorem 2 (Multi-hop high-pass filtering under heterophily)**.** *Consider an undirected regular graph in domain a with positive degree and no self-loops. Let $\mathbf{v}_a \in \{-1, 1\}^{n_a}$ be the class signal, and suppose that every node has the same fraction $\vartheta_a < 1/2$ of same-class neighbors. For every $j \geq 1$,*

$$\|\mathbf{P}_{a,H}^j \mathbf{v}_a\|_2 = (1-\vartheta_a)^j \|\mathbf{v}_a\|_2 > \vartheta_a^j \|\mathbf{v}_a\|_2 = \|\mathbf{P}_{a,L}^j \mathbf{v}_a\|_2. \tag{4}$$

Let $\mathbf{e}_a$ be an interference signal supported on eigenvectors of $\tilde{\mathbf{A}}_a$ with eigenvalues at least $\nu_{a,0} > 2\vartheta_a - 1$, with $\nu_{a,0} \leq 1$. This places the interference at lower graph frequencies than the class signal. Then,

$$\frac{\|\mathbf{P}_{a,H}^j \mathbf{e}_a\|_2}{\|\mathbf{P}_{a,H}^j \mathbf{v}_a\|_2} \leq \left(\frac{1-\nu_{a,0}}{2(1-\vartheta_a)}\right)^j \frac{\|\mathbf{e}_a\|_2}{\|\mathbf{v}_a\|_2}. \tag{5}$$

Since $(1-\nu_{a,0})/[2(1-\vartheta_a)] \in [0, 1)$, the relative-interference bound decays with propagation order.

The proof is provided in Appendix A.6.2. Eq. (4) indicates that relying only on low-pass filtering can strongly attenuate the class signal in this heterophilic setting, whereas a high-pass view retains more of it. When interference occupies lower graph frequencies, repeated high-pass filtering attenuates it faster than the class signal as reflected in Eq. (5). This motivates multi-hop high-pass views, while identity and low-pass views preserve information that may not follow this spectral pattern in real graphs.

4.2 Cross Domain Prototype Adaptation

Deep Embedded Clustering (DEC) learns representations by matching soft cluster assignments to a sharpened target distribution (Xie et al., 2016). We retain the DEC clustering objective but adapt prototype construction to open-set graph domain adaptation. For each known class c, we compute

the source prototype $\boldsymbol{\mu}_c^s$ as the mean representation of source nodes with label c. Then we obtain target assignments to these source prototypes using a Student-t kernel:

$$q_{t,i,c}^s = \frac{(1 + \|\mathbf{z}_{t,i} - \boldsymbol{\mu}_c^s\|_2^2)^{-1}}{\sum_{b=0}^{C-1}(1 + \|\mathbf{z}_{t,i} - \boldsymbol{\mu}_b^s\|_2^2)^{-1}}. \tag{6}$$

We assign weight $w_{i,c} = p_{t,i,c}q_{t,i,c}^s$, giving greater influence to nodes supported by both the classifier and source-prototype assignments. Thus we define the target prototype $\boldsymbol{\mu}_c^t = (\sum_i w_{i,c}\mathbf{z}_{t,i})/(\sum_i w_{i,c})$. However, target predictions may be inaccurate under domain shift and unknown nodes may receive weights for known classes. To control its contribution, we define a class-wise reliability score $r_c = (\sum_i w_{i,c})/(\sum_i p_{t,i,c})$. This score measures the average support from source-prototype assignments, weighted by classifier probabilities. The final known-class prototype combines source and target information from the insight of Theorem 3: $\boldsymbol{\mu}_c = (\boldsymbol{\mu}_c^s + r_c\boldsymbol{\mu}_c^t)/(1+r_c)$. To accommodate the open-set setting, we initialize the unknown prototype $\boldsymbol{\mu}_C$ as the mean representation of the top k target nodes ranked by predicted unknown probability. Using the resulting $C+1$ prototypes, we compute assignments $q_{t,i,c}$ with the same Student-t kernel, normalized over $c = 0, \ldots, C$. Following DEC, the sharpened target distribution is

$$\widehat{q}_{t,i,c} = \frac{q_{t,i,c}^2 / \sum_m q_{t,m,c}}{\sum_{b=0}^{C} q_{t,i,b}^2 / \sum_m q_{t,m,b}}, \tag{7}$$

and the clustering loss is

$$\mathcal{L}_{\text{DEC}} = \frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{c=0}^{C}\widehat{q}_{t,i,c}\log\frac{\widehat{q}_{t,i,c}}{q_{t,i,c}}. \tag{8}$$

Theorem 3 (Source-constrained prototype adaptation)**.** *For a known class c with positive target weight, consider $\|\boldsymbol{\mu} - \boldsymbol{\mu}_c^s\|_2^2 + r_c\|\boldsymbol{\mu} - \boldsymbol{\mu}_c^t\|_2^2$, where the first term anchors the prototype to the source class center and the second encourages target adaptation. Its unique minimizer is the adapted prototype $\boldsymbol{\mu}_c$ defined above. Let $\boldsymbol{m}_{t,c} = \mathbb{E}_t[\mathbf{z} \mid Y = c]$ be the true target class center. Define $b_c = \|\boldsymbol{\mu}_c^s - \boldsymbol{m}_{t,c}\|_2$ and $e_c = \|\boldsymbol{\mu}_c^t - \boldsymbol{m}_{t,c}\|_2$. Then,*

$$\|\boldsymbol{\mu}_c - \boldsymbol{m}_{t,c}\|_2 \le (b_c + r_c e_c)/(1 + r_c). \tag{9}$$

The proof is provided in Appendix A.6.3. Eq. (9) shows that, when $r_c > 0$ and the target estimate is more accurate than the source prototype ($e_c < b_c$), the update moves the prototype closer to the true target class center.

4.3 Consensus-Guided Learning

Next, we select target pseudo labels through agreement between the classifier and prototype-based clustering, whose predictions are $\widehat{y}_{t,i}^{\text{cls}} = \arg\max_c p_{t,i,c}$ and $\widehat{y}_{t,i}^{\text{DEC}} = \arg\max_c q_{t,i,c}$. Nodes with matching predictions receive the agreed pseudo-label $\widetilde{y}_{t,i}$. The target-domain classification loss is

$$\mathcal{L}_{\text{con}} = \lambda_k \,\text{CE}(\mathbf{P}_{t,k}, \widetilde{\mathbf{y}}_{t,k}) + \lambda_u \,\text{CE}(\mathbf{P}_{t,u}, \widetilde{\mathbf{y}}_{t,u}), \tag{10}$$

where $\mathbf{P}_{t,k}$ and $\mathbf{P}_{t,u}$ contain classifier probabilities for consensus-selected known and unknown nodes, and CE denotes mean cross-entropy. The same consensus guides domain alignment by excluding nodes jointly predicted as unknown. We align the identity, low-pass, and high-pass channels separately, concatenating representations across propagation orders as $\mathbf{H}_{a,r} = \mathbf{H}_{a,r,1}\|\cdots\|\mathbf{H}_{a,r,J}$ for $r \in \{L, H\}$. Following Gretton et al. (2012), we use maximum mean discrepancy (MMD) with a Gaussian radial basis function (RBF) kernel and average the three channel-wise losses:

$$\mathcal{L}_{\text{align}} = \frac{1}{3}\sum_{r\in\{I,L,H\}}\text{MMD}_{\text{RBF}}^2\left(\mathbf{H}_{s,r}, \mathbf{H}_{t,r}^{\text{keep}}\right), \tag{11}$$

where $\mathbf{H}_{t,r}^{\text{keep}}$ contains the target representations retained for alignment.

4.4 Training Procedure

We train the framework in two stages. To learn discriminative source representations and initialize unknown-class prediction before target adaptation, the first stage uses source supervision. We define $\mathcal{L}_{\text{src}} = \text{CE}(\mathbf{P}_s, \mathbf{y}_s)$ and introduce an auxiliary loss $\mathcal{L}_{\text{leaveout}}$. For this auxiliary objective, we mask each source node's true-class logit and apply cross-entropy with the unknown label C. The first-stage objective is

$$\mathcal{L}_{\text{warm}} = \mathcal{L}_{\text{src}} + \lambda_{\text{warm}} \mathcal{L}_{\text{leaveout}}. \tag{12}$$

The second stage adapts the model to the target graph while preserving source class information. After initializing the unknown prototype as described above, we jointly optimize

$$\mathcal{L} = \mathcal{L}_{\text{con}} + \lambda_s \mathcal{L}_{\text{src}} + \lambda_{\text{DEC}} \mathcal{L}_{\text{DEC}} + \lambda_{\text{align}} \mathcal{L}_{\text{align}}, \tag{13}$$

During adaptation, the encoder, classifier, and unknown prototype are optimized at each iteration, while known-class prototypes and consensus pseudo labels are recomputed. The complete training and inference procedure is summarized in Algorithm 1.

5 Experiments

Datasets. We evaluate HODA on four widely used graph benchmarks: Blog, WebKB, and MAG for heterophilic settings, and Citation for a homophilic setting. Dataset statistics are provided in Table 4. **Blog** consists of two social networks, Blog1 (B1) and Blog2 (B2), derived from BlogCatalog (Shen et al., 2020). **WebKB** contains webpage networks from Cornell (CO), Texas (TX), and Wisconsin (WI). We use the processed graphs released with Geom-GCN (Pei et al., 2020), whose numerical label mappings are inconsistent across universities. We recover semantic labels by matching node features to the original WebKB content records (Craven et al., 1998) and apply a shared encoding. Table 5 in Appendix A.2 reports the original and corrected mappings. **MAG** comprises six country-specific graphs derived from the Microsoft Academic Graph (Hu et al., 2020). **Citation** contains three citation networks, ACMv9 (A), Citationv1 (C), and DBLPv7 (D), extracted from ArnetMiner (Shen et al., 2020).

Baselines. We compare HODA with three groups of baselines. For heterophilic graph learning, we include H2GCN (Zhu et al., 2020), which incorporates higher order neighborhoods; ACM (Luan et al., 2022) and HeroFilter (Zhang et al., 2026), which adaptively combine complementary graph channels and spectral responses; CMGNN (Zheng et al., 2026), which models class compatibility to guide neighborhood aggregation; and HIMP (Wang et al., 2025), which promotes homophily in the learned embedding space. For closed-set GDA, we consider Pair-Align (Liu et al., 2024), which reweights edges to mitigate structural shifts; ADAlign (Chen et al., 2026a), which adaptively matches distributions through neural characteristic functions; DFT (Tai et al., 2026), which uses feature decorrelation to mitigate local dependencies; DGSDA (Yang et al., 2025), which disentangles attribute and topology alignment through spectral filtering; DiffGDA (Chen et al., 2026b), which models cross domain transitions through guided diffusion; and RSGDA (Fang et al., 2026), which reconstructs homophilic and heterophilic graph views and aligns them separately. For open-set GDA, we include UAGA (Shen et al., 2025), which aligns known target nodes with the source while pushing unknown nodes away; and GraphRTA (Zhang & He, 2026), which uses graph and model reprogramming to improve known-unknown separation. We search the reported hyperparameter ranges for each baseline and equip closed-set methods with unknown rejection based on maximum softmax probability (MSP) (Hendrycks & Gimpel, 2016) or predictive entropy.

Implementation and evaluation. We implement HODA in PyTorch using Adam. The search ranges cover propagation orders 1-4, MLP depths 1-2, hidden dimensions 64-256, learning rates $[10^{-4}, 10^{-3}]$ for both training stages, loss weights $[0.1, 2]$ for λ_{warm}, λ_s, λ_k, and λ_u, $[10^{-3}, 1]$ for λ_{DEC} and λ_{align}, and unknown-candidate ratios $[0.1, 0.5]$. Following prior settings (Wang et al., 2024; Shen et al., 2025; Zhang & He, 2026), the last two classes in Blog, WebKB, and Citation are designated unknown. For MAG, the final label aggregates categories outside the retained top 19 classes (Liu et al., 2024). The unknown set contains either this aggregate class alone or the last five labels. Unknown-class nodes and their incident edges are removed from the source graph; all target nodes are retained. Evaluation uses ACC over all target nodes, $\text{HS} = 2A_kA_u/(A_k + A_u)$, and AUC for unknown detection with unknown nodes as positives (Saito et al., 2018; Fu et al., 2020), where A_k and A_u are node-level known and unknown class accuracies. Results are reported as mean and standard deviation over five seeds.

Table 1: Open-set node classification on **Blog and WebKB**. Results are mean ± standard deviation (%). Bold and underline indicate the best and second-best means.

Task	Metric	Heterophilic GNNs					Closed-set DA						Open-set DA		HODA
		ACM	CMGNN	HeroFilter	H2GCN	HIMP	ADAlign	DiffGDA	Pair-Align	DGSDA	DFT	RSGDA	UAGA	GraphRTA	
B1 → B2	ACC	52.18±2.26	52.52±4.20	44.26±3.38	51.52±0.60	45.24±1.22	32.89±6.07	25.64±1.06	38.25±4.88	52.77±2.11	50.65±5.53	40.82±0.99	26.56±4.25	26.55±1.67	**60.46±0.38**
	HS	52.46±1.87	51.61±7.20	44.96±3.39	52.74±0.57	44.42±2.89	29.45±3.71	25.67±1.09	38.25±4.90	53.69±2.10	47.38±9.81	38.66±0.53	17.20±12.19	26.33±2.33	**61.94±0.39**
	AUC	73.22±2.19	67.75±3.72	59.25±3.87	64.22±0.56	59.96±1.74	49.28±6.51	45.30±2.28	53.50±2.58	63.42±2.64	63.13±5.07	58.26±0.74	60.44±7.14	55.56±0.55	**76.40±0.46**
B2 → B1	ACC	58.55±0.85	51.10±2.88	46.86±3.43	55.71±0.28	47.45±2.01	33.80±4.99	27.49±3.70	50.87±2.65	51.00±0.34	52.40±6.17	39.97±0.93	23.86±5.93	27.19±3.84	**58.83±0.77**
	HS	60.04±0.86	52.22±2.90	44.98±6.34	56.95±0.41	48.64±1.96	29.14±4.25	24.93±4.48	51.43±2.68	51.05±0.60	52.01±6.02	40.45±0.65	17.54±11.32	17.25±14.31	**60.48±0.83**
	AUC	71.41±3.21	64.27±1.55	61.01±3.84	67.99±0.23	65.97±2.15	53.46±4.51	48.58±2.06	65.65±2.62	63.33±0.40	66.45±5.27	58.31±0.48	59.48±7.98	48.41±2.81	**77.22±0.81**
TX → CO	ACC	75.19±1.01	68.63±2.92	58.80±7.35	73.01±0.74	73.11±0.64	35.63±2.55	44.37±14.21	56.72±10.38	70.93±1.60	57.05±5.99	54.43±10.81	48.42±2.96	48.96±3.57	**77.38±3.33**
	HS	72.61±0.84	64.10±0.99	55.50±10.71	72.24±0.50	69.61±2.02	33.94±2.49	13.90±11.88	49.39±8.55	69.83±1.30	56.45±5.80	50.21±7.14	42.33±0.83	40.53±8.07	**76.10±2.89**
	AUC	85.19±0.29	73.35±1.92	76.98±3.24	**86.48±0.63**	80.59±0.54	58.44±2.11	49.40±1.95	62.95±1.65	85.83±1.16	68.83±8.36	64.66±3.12	52.93±1.15	47.59±4.44	85.77±2.15
TX → WI	ACC	64.46±6.45	72.27±2.87	71.16±1.88	74.66±0.20	70.28±1.30	42.39±2.65	40.96±6.71	56.33±7.87	73.86±1.30	56.02±3.88	64.46±6.35	50.92±6.56	52.91±8.66	**80.16±2.16**
	HS	62.19±5.37	69.54±2.91	69.13±2.15	71.84±0.13	67.23±1.39	39.80±4.63	35.06±9.38	23.25±24.17	73.16±1.07	55.72±4.05	63.58±5.46	42.39±14.95	48.79±7.32	**77.85±3.56**
	AUC	70.08±4.36	73.65±1.50	76.58±1.72	80.61±0.02	71.16±0.60	58.28±1.59	59.45±2.57	50.45±20.44	83.72±0.64	65.08±4.59	77.83±2.92	63.35±2.88	58.33±5.52	**86.79±1.50**
CO → TX	ACC	77.27±2.06	66.56±5.26	74.97±3.81	77.60±0.60	75.08±2.00	27.65±13.00	39.67±4.92	69.07±1.91	78.36±1.01	71.04±4.69	73.88±1.78	65.90±4.34	58.03±5.32	**80.98±1.36**
	HS	69.54±3.08	66.54±4.70	65.29±3.92	71.64±2.18	66.22±4.64	15.75±17.90	38.04±3.72	62.74±2.80	69.38±2.53	58.01±4.34	60.77±1.01	48.97±11.72	51.43±3.06	**72.25±4.18**
	AUC	81.63±1.82	70.87±3.47	73.84±5.79	80.06±0.97	79.26±2.67	54.41±8.00	61.86±2.53	71.01±2.24	81.51±1.97	70.07±3.70	65.75±4.07	79.63±3.69	68.54±3.13	**83.59±2.78**
CO → WI	ACC	78.25±1.35	82.07±0.56	72.43±1.41	80.48±1.26	74.98±2.10	59.84±2.41	37.05±8.33	74.90±1.40	80.48±1.63	65.10±8.48	71.00±2.46	50.28±6.56	59.20±3.44	**85.02±1.52**
	HS	77.40±1.57	76.87±0.78	69.47±3.80	77.92±1.12	73.17±1.32	52.42±4.55	24.78±20.59	67.84±4.80	80.08±1.43	59.54±7.91	66.77±2.05	48.30±8.67	57.72±2.62	**82.47±1.66**
	AUC	84.65±0.58	86.90±0.46	77.65±1.26	85.12±0.65	79.73±1.44	71.82±2.00	60.76±1.41	76.88±0.90	86.59±0.76	72.25±7.15	66.89±1.72	59.99±1.16	63.10±1.95	**92.19±0.62**
WI → TX	ACC	80.77±3.56	82.40±1.44	79.23±2.21	75.85±1.81	62.08±1.45	57.81±6.06	46.01±6.98	75.85±1.64	77.38±1.32	69.51±7.54	76.17±2.20	49.18±11.38	51.91±13.01	**82.95±1.81**
	HS	78.16±2.21	79.10±2.18	73.08±2.83	73.77±1.49	60.35±0.67	48.74±4.58	36.92±7.13	70.53±3.41	62.72±1.35	63.68±4.54	62.45±1.28	46.40±8.16	46.25±11.36	**79.60±1.29**
	AUC	86.78±1.22	83.08±1.20	83.34±2.34	80.87±0.48	71.63±0.96	70.73±4.67	62.99±3.72	76.33±2.94	71.09±1.08	76.50±3.03	78.52±2.20	75.98±2.81	54.44±5.24	**87.59±1.31**
WI → CO	ACC	73.01±0.89	72.57±3.14	67.43±3.70	70.60±0.41	65.57±1.09	52.90±0.94	28.20±6.15	54.97±4.41	61.42±3.53	66.67±2.72	63.93±1.92	54.21±4.62	52.68±2.31	**75.19±4.01**
	HS	71.85±0.80	72.17±2.69	65.49±3.85	70.29±0.15	56.77±0.94	37.90±1.49	19.11±15.66	53.74±4.55	58.59±4.13	59.85±4.84	55.97±1.62	44.49±7.03	36.88±16.68	**73.61±4.71**
	AUC	84.79±0.87	**85.03±1.33**	75.52±3.17	81.24±0.23	64.29±1.17	59.93±2.87	49.15±0.95	64.69±5.32	70.08±1.41	74.28±3.25	57.27±0.94	64.66±3.19	53.29±2.58	81.17±1.82

6 Overall Performance

Tables 1 and 2 summarize performance on the heterophilic benchmarks. Results with one unknown class on MAG are reported in Table 7. All reported gains are absolute differences between percentage-valued metrics. Across the 18 transfer tasks, HODA achieves the best HS on all tasks and the best ACC on 17 tasks. The only exception is CN→RU, where RSGDA attains higher ACC with a standard deviation above 12% and an HS below 12%. Compared with the baseline achieving the highest average for each metric, HODA improves ACC and HS by 4.28% and 4.96% on Blog and by 4.91% and 4.03% on WebKB. On MAG, the corresponding gains are 4.16% and 2.27% with one unknown class and 4.74% and 3.12% with five unknown classes.

The baseline categories exhibit complementary weaknesses that depend on the degree of heterophily. Open-set DA methods are designed to identify unknown nodes but assume that neighboring nodes share similar labels. They therefore degrade sharply on the strongly heterophilic Blog and WebKB graphs. Heterophilic GNNs handle heterophilic structure and are the strongest competitors on Blog and WebKB. However, they neither align the two domains nor model unknown classes, which limits their HS under open-set shifts. Closed-set DA methods align the domains without accounting for unknown classes, and only DGSDA is consistently competitive among them. In contrast, HODA addresses heterophily and unknown classes jointly and therefore performs best across all benchmarks. The improvements also extend to the homophilic Citation benchmark (Table 6), where HODA exceeds the runner-up average by 5.52% in ACC and 3.54% in HS. These results indicate that HODA is effective across different graph structures and open-set settings.

7 Ablation Study

Component analysis. Table 3 compares HODA with three variants on Blog and WebKB. (1) w/o MSE replaces multiscale spectral encoding with the original attributes and one-hop normalized adjacency propagation. (2) w/o PA uses only source known-class prototypes. (3) w/o CF directly uses classifier predictions as target pseudo labels without consensus filtering. Removing multiscale spectral encoding or prototype adaptation reduces HS across all tasks, with the largest drops of 21.29% and 8.55%, respectively, on TX→WI. Without consensus filtering, HS drops by 45.90% on TX→CO, highlighting the importance of screening target pseudo labels. Additional results in Table 8 support the contributions of high-pass views and task dependent benefits of higher propagation orders. Adaptive prototype mixing and consensus selection also achieve higher mean ACC and HS than equal weight mixing and confidence filtering, respectively.

Representation visualization. Figure 3 visualizes target node representations on B1→B2 using t-SNE, comparing HODA with the best-performing baseline from each method category. Red indicates unknown nodes, while other colors represent known classes. H2GCN and DGSDA show considerable overlap between classes. Although GraphRTA produces several separated groups, nodes from different classes remain mixed within them. HODA exhibits more distinct concentrations of

Table 2: Open-set node classification on **MAG**. The last five classes are treated as unknown. Results are mean ± standard deviation (%). Bold and underline indicate the best and second-best means.

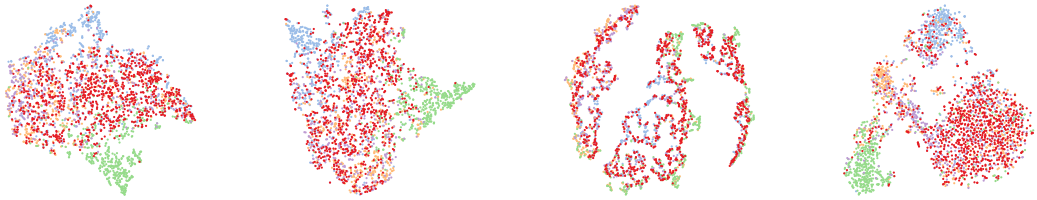
Task	Metric	Heterophilic GNNs					Closed-set DA						Open-set DA		HODA
		ACM	CMGNN	HeroFilter	H2GCN	HIMP	ADAlign	DiffGDA	Pair-Align	DGSDA	DFT	RSGDA	UAGA	GraphRTA	
CN → DE	ACC	39.62±0.32	40.96±1.99	34.95±0.36	41.37±0.19	39.26±0.37	41.73±0.21	38.61±0.33	37.66±0.79	39.79±0.28	34.54±1.45	30.05±4.32	42.10±0.98	36.71±2.72	**44.23±0.20**
	HS	36.66±0.08	32.48±2.70	33.71±0.42	37.60±0.12	35.79±0.28	34.64±0.24	33.60±0.26	34.84±0.67	32.69±0.29	32.17±0.98	12.25±1.57	33.48±1.42	34.82±1.65	**41.45±0.15**
	AUC	60.39±0.64	69.94±1.75	51.88±0.50	62.71±0.14	59.60±0.61	62.86±0.22	58.27±0.53	49.44±1.05	64.79±0.13	51.39±3.64	45.50±4.39	61.90±5.70	72.23±1.13	**73.74±0.32**
CN → FR	ACC	35.60±0.93	39.57±2.20	31.46±0.41	39.36±0.09	34.24±0.44	36.81±0.51	34.38±0.77	34.14±1.97	34.10±0.23	32.00±0.65	28.95±5.56	35.76±1.65	37.31±2.12	**43.63±0.14**
	HS	32.75±0.47	28.58±0.98	29.51±0.36	33.11±0.08	32.12±0.19	32.28±0.22	28.98±0.24	26.31±2.64	29.54±0.18	27.89±0.23	12.32±1.75	29.97±1.62	33.06±0.75	**38.30±0.29**
	AUC	60.15±0.78	66.36±4.51	46.88±0.75	61.83±0.11	58.17±0.67	59.58±0.73	54.82±1.21	45.89±4.10	61.89±0.14	49.37±2.26	46.05±8.21	50.39±8.17	70.69±1.95	**73.64±0.72**
CN → JP	ACC	48.43±0.45	49.31±0.63	47.89±2.41	49.02±0.10	49.21±0.95	47.18±0.56	46.31±0.63	43.53±3.78	46.65±0.34	43.05±0.87	40.49±3.77	46.78±0.69	50.50±2.78	**54.75±0.86**
	HS	45.26±0.15	44.15±0.26	40.14±1.98	46.20±0.05	44.29±0.99	44.79±0.18	42.84±0.36	37.78±2.65	43.28±0.14	40.00±0.52	17.78±5.51	41.82±0.31	40.73±0.76	**46.47±0.94**
	AUC	60.42±0.29	63.45±0.57	58.73±2.08	62.00±0.05	62.04±0.89	60.73±0.36	57.17±0.58	50.93±3.08	58.72±0.19	51.08±1.58	47.24±2.37	59.53±0.74	62.04±0.63	**68.41±1.00**
CN → RU	ACC	45.90±1.50	47.15±2.46	42.33±1.64	45.48±0.37	46.75±0.85	43.81±0.47	46.00±1.33	43.69±3.24	50.06±1.91	46.61±2.23	**59.02±12.40**	44.56±4.32	43.76±5.86	52.10±1.31
	HS	39.80±0.57	35.34±0.54	37.55±0.52	40.73±0.10	39.97±0.52	39.19±0.11	37.52±0.46	37.25±1.49	39.22±0.68	36.72±1.34	11.89±2.79	36.01±1.68	37.73±0.53	**45.79±0.74**
	AUC	54.54±0.76	59.18±1.10	49.69±0.79	57.50±0.17	55.16±0.69	57.15±0.28	53.48±0.29	50.84±3.20	54.74±0.87	50.69±0.99	50.47±3.94	66.31±0.67	63.27±0.70	**73.16±0.24**
CN → US	ACC	51.02±0.19	50.19±0.93	42.88±2.65	50.23±0.05	49.13±0.96	48.84±0.21	46.52±0.53	42.28±3.79	47.60±0.22	44.08±1.16	27.79±4.32	47.15±3.05	48.78±2.02	**52.24±0.44**
	HS	48.47±0.19	47.20±0.27	38.68±7.07	48.81±0.03	47.23±0.91	46.72±0.15	45.15±0.74	38.51±5.73	45.09±0.14	43.17±1.06	12.34±2.53	45.35±2.46	42.98±1.54	**50.79±0.57**
	AUC	62.05±0.44	67.40±0.66	54.20±4.13	63.56±0.05	65.87±1.08	64.05±0.27	57.90±1.64	56.16±3.84	62.19±0.23	54.49±1.54	46.58±5.85	66.48±3.88	66.17±1.15	**73.60±0.34**
US → CN	ACC	50.00±0.74	45.45±2.34	40.76±1.05	48.54±0.02	47.79±0.55	46.05±0.65	50.53±0.54	46.77±2.39	44.13±0.33	39.91±5.55	26.90±9.26	44.29±1.12	54.83±4.05	**56.85±2.23**
	HS	41.64±0.66	38.57±1.81	39.76±0.74	43.18±0.03	40.81±0.11	42.43±0.20	41.61±0.32	37.52±1.51	37.47±0.42	35.68±2.56	19.83±7.00	40.11±0.47	31.86±3.70	**43.80±1.40**
	AUC	60.31±0.34	58.05±1.05	54.31±0.61	60.25±0.02	55.59±0.24	60.30±0.17	56.41±0.20	53.61±0.81	55.48±0.54	51.55±1.70	52.07±3.80	54.55±0.63	58.09±1.06	**68.96±0.45**
US → DE	ACC	41.61±0.52	44.17±1.69	38.66±0.55	42.41±0.02	39.85±0.41	36.74±0.20	38.38±0.90	40.61±1.29	41.42±0.12	44.62±1.96	26.92±4.49	48.43±1.11	41.02±2.70	**49.89±1.21**
	HS	38.81±0.23	37.17±1.81	36.27±0.58	38.99±0.02	37.31±0.14	36.05±0.17	35.98±0.54	39.81±1.27	36.31±0.15	37.65±1.33	20.55±2.08	36.91±1.09	38.98±1.37	**42.48±0.67**
	AUC	58.29±1.38	68.35±0.74	57.75±0.52	65.55±0.02	61.22±0.60	72.25±0.20	68.03±1.36	53.71±1.26	72.35±0.08	73.36±1.13	69.10±3.15	74.21±1.40	72.85±2.65	**75.23±0.32**
US → FR	ACC	40.04±0.93	36.95±0.84	36.33±0.74	40.07±0.09	38.56±1.23	33.44±0.20	34.67±0.69	39.86±1.61	37.51±0.38	44.18±1.82	29.20±6.04	38.61±2.42	48.63±1.91	**50.03±0.53**
	HS	35.09±0.33	31.59±0.32	31.63±0.25	34.72±0.06	33.48±0.72	31.75±0.10	30.06±0.44	37.58±1.77	32.45±0.20	32.98±0.99	19.80±2.36	34.53±1.39	26.88±1.88	**38.62±0.91**
	AUC	67.85±1.71	71.79±0.25	55.89±0.82	63.72±0.03	71.21±1.66	69.33±0.23	63.44±0.65	52.07±2.35	65.40±0.52	73.25±1.34	68.63±3.92	64.75±5.01	73.25±0.39	**74.22±0.89**
US → JP	ACC	48.28±0.20	48.15±1.55	45.95±0.61	52.12±0.11	51.28±0.99	48.41±1.06	49.57±0.22	43.64±2.29	45.90±0.33	45.49±2.28	31.33±8.39	52.74±1.12	52.97±0.92	**55.75±0.91**
	HS	45.98±0.17	44.95±0.85	41.85±0.49	46.62±0.04	44.67±0.46	45.90±0.47	43.96±0.18	42.32±1.53	44.14±0.22	44.01±1.51	28.07±7.10	45.40±0.26	43.26±0.63	**47.07±0.51**
	AUC	65.10±0.24	67.22±1.28	52.05±0.72	64.87±0.11	61.61±1.16	62.93±1.17	61.21±0.49	56.59±1.72	59.08±0.37	63.48±1.76	58.35±5.59	65.89±0.67	64.45±0.35	**69.41±0.25**
US → RU	ACC	49.59±0.05	48.93±0.99	47.97±0.27	45.47±0.03	46.73±1.64	41.12±0.48	50.62±0.48	50.37±1.27	46.79±1.11	50.15±2.85	26.33±6.04	52.85±2.33	52.08±7.09	**54.50±0.95**
	HS	39.64±0.29	36.55±0.95	37.93±0.32	38.77±0.06	38.45±0.52	37.02±0.14	37.28±0.37	42.89±1.19	36.22±0.17	37.42±0.63	14.57±2.54	40.10±1.77	38.97±1.02	**45.19±0.61**
	AUC	54.96±0.06	60.18±0.44	50.58±0.46	58.20±0.02	51.56±0.35	58.73±0.26	54.20±0.36	57.06±2.07	57.13±0.30	52.60±0.74	53.48±3.18	68.43±1.26	64.86±0.96	**73.56±0.18**

known classes and a region dominated by unknown nodes, while some overlap remains. These observations qualitatively support its ability to preserve class information while improving known-unknown separation under heterophilic domain shift.

Parameter sensitivity. Figure 4 in Appendix A.4 examines nine hyperparameters on Blog and WebKB. Removing classification loss terms by setting their weights to zero can substantially reduce ACC. With appropriately chosen nonzero weights, performance is relatively stable over a range of nearby settings, although sensitivity varies across parameters and tasks.

Table 3: Ablation study on **Blog and WebKB**. Results are mean ± standard deviation (%) over five seeds. Bold indicates the best mean.

Variant	B1→B2		B2→B1		TX→CO		TX→WI		CO→WI	
	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS
w/o MSE	59.27±0.80	60.63±0.83	57.31±0.61	58.64±0.87	58.14±7.98	56.35±6.29	63.82±13.56	56.56±14.62	79.76±1.43	75.37±2.68
w/o PA	60.08±0.38	61.37±0.32	58.25±0.57	59.78±0.63	74.64±2.36	70.31±2.61	77.61±3.63	69.30±6.04	81.20±1.39	74.52±2.54
w/o CF	58.87±0.54	58.86±0.79	54.76±1.06	51.71±1.75	62.08±1.07	30.20±3.73	72.11±1.43	56.31±2.26	77.53±0.32	66.60±0.51
HODA	**60.46±0.38**	**61.94±0.39**	**58.83±0.77**	**60.48±0.83**	**77.38±3.33**	**76.10±2.89**	**80.16±2.16**	**77.85±3.56**	**85.02±1.52**	**82.47±1.66**



(a) H2GCN (b) DGSDA (c) GraphRTA (d) HODA

Figure 3: t-SNE visualization of node representations on B1→B2.

8 Conclusion

We propose HODA, a unified framework for open-set graph domain adaptation under heterophily. The framework combines multiscale spectral encoding, source prototype adaptation, and classifier–clustering consensus to support known-class classification and unknown-class recognition. We also correct cross domain label inconsistencies in the processed WebKB graphs to enable semantically consistent evaluation. Experiments across multiple datasets and transfer directions show competitive performance.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation, sourcing and searching for information, editing the manuscript to improve readability, suggesting ideas to support the authors in developing and writing proofs, and refining existing code. We did not use generative AI tools for any other tasks listed in the disclosure policy. The first author reviewed all AI-assisted content, including the text, proofs, and code, and confirmed its correctness and appropriateness. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The code and datasets are available at the anonymous repository https://anonymous.4open.science/r/ICLR2027_HODA/, and the code will be made publicly available upon acceptance. The methodology and Algorithm 1 describe the model and training procedure. Dataset statistics and corrected WebKB label mappings are provided in Tables 4 and 5. Complete proofs are included in Appendix A.6.

REFERENCES

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.

Wei Chen, Xingyu Guo, Shuang Li, Zhao Zhang, Yan Zhong, Fuzhen Zhuang, et al. Learning adaptive distribution alignment with neural characteristic function for graph domain adaptation. *arXiv preprint arXiv:2602.10489*, 2026a.

Wei Chen, Xingyu Guo, Shuang Li, Yan Zhong, Zhao Zhang, Fuzhen Zhuang, Hongrui Liu, Libang Zhang, Guo Ye, and Huimei He. Learning structure-semantic evolution trajectories for graph domain adaptation. *arXiv preprint arXiv:2602.10506*, 2026b.

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, Tom Mitchell, K. Nigam, and S. Slattery. Learning to extract symbolic knowledge from the world wide web. In *Proceedings of 15th National Conference on Artificial Intelligence (AAAI '98)*, pp. 509 – 516, July 1998.

Vijay Prakash Dwivedi, Chaitanya K Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24(43):1–48, 2023.

Ruiyi Fang, Bingheng Li, Jingyu Zhao, Ruizhi Pu, Qiuhao Zeng, Gezheng Xu, Charles Ling, and Boyu Wang. Homophily enhanced graph domain adaptation. *arXiv preprint arXiv:2505.20089*, 2025.

Ruiyi Fang, Shuo Wang, Ruizhi Pu, Qiuhao Zeng, Hao Zheng, Ziyan Wang, Jiale Cai, Zhimin Mei, Song Tang, Charles Ling, et al. Graph domain adaptation via homophily-agnostic reconstructing structure. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 21047–21055, 2026.

Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Learning to detect open classes for universal domain adaptation. In *European conference on computer vision*, pp. 567–583. Springer, 2020.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The journal of machine learning research*, 13:723–773, 2012.

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

Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2016.

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

Shikun Liu, Deyu Zou, Han Zhao, and Pan Li. Pairwise alignment improves graph domain adaptation. *arXiv preprint arXiv:2403.01092*, 2024.

Sitao Luan, Chenqing Hua, Qincheng Lu, Jiaqi Zhu, Mingde Zhao, Shuyuan Zhang, Xiao-Wen Chang, and Doina Precup. Revisiting heterophily for graph neural networks. *Advances in neural information processing systems*, 35:1362–1375, 2022.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. *arXiv preprint arXiv:2002.05287*, 2020.

Ruizhong Qiu, Ting-Wei Li, Gaotang Li, and Hanghang Tong. Graph homophily booster: Reimagining the role of discrete features in heterophilic graph learning. *arXiv preprint arXiv:2602.07256*, 2026.

Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *European conference on computer vision*, pp. 156–171. Springer, 2018.

Xiao Shen, Quanyu Dai, Fu-lai Chung, Wei Lu, and Kup-Sze Choi. Adversarial deep network embedding for cross-network node classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 2991–2999, 2020.

Xiao Shen, Zhihao Chen, Shirui Pan, Shuang Zhou, Laurence T Yang, and Xi Zhou. Open-set cross-network node classification via unknown-excluded adversarial graph domain alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20398–20408, 2025.

Xinwei Tai, Dongmian Zou, and Hongfei Wang. Enhancing node-level graph domain adaptation by alleviating local dependency. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 1366–1377, 2026.

Haishan Wang, Arno Solin, and Vikas Garg. Heterophily-informed message passing. *arXiv preprint arXiv:2504.19785*, 2025.

Yu Wang, Ronghang Zhu, Pengsheng Ji, and Sheng Li. Open-set graph domain adaptation via separate domain alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 9142–9150, 2024.

Man Wu, Shirui Pan, Chuan Zhou, Xiaojun Chang, and Xingquan Zhu. Unsupervised domain adaptive graph convolutional networks. In *Proceedings of the web conference 2020*, pp. 1457–1467, 2020.

Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pp. 478–487. PMLR, 2016.

Liang Yang, Xin Chen, Jiaming Zhuo, Di Jin, Chuan Wang, Xiaochun Cao, Zhen Wang, and Yuanfang Guo. Disentangled graph spectral domain adaptation. In *Forty-second International Conference on Machine Learning*, 2025.

Shuaicheng Zhang, Haohui Wang, Junhong Lin, Xiaojie Guo, Yada Zhu, Si Zhang, Dongqi Fu, and Dawei Zhou. Herofilter: Adaptive spectral graph filter for varying heterophilic relations. *Advances in Neural Information Processing Systems*, 38:64537–64559, 2026.

Zhen Zhang and Bingsheng He. Towards unsupervised open-set graph domain adaptation via dual reprogramming. *Advances in Neural Information Processing Systems*, 38:53320–53343, 2026.

Zhuonan Zheng, Yuanchen Bei, Zhiyao Zhou, Sheng Zhou, Yao Ma, Ming Gu, Hongjia Xu, Jiawei Chen, and Jiajun Bu. Understanding and enhancing message passing on heterophilic graphs via compatibility matrix. *Advances in Neural Information Processing Systems*, 38:54386–54418, 2026.

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in neural information processing systems*, 33:7793–7804, 2020.

A APPENDIX

A.1 DATASET STATISTICS

Table 4: Dataset statistics before open-set splits.

Dataset	#Nodes	#Edges	#Features	#Classes
ACMv9	9,360	15,556	6,775	5
Citationv1	8,935	15,098	6,775	5
DBLPv7	5,484	8,117	6,775	5
Blog1	2,300	33,471	8,189	6
Blog2	2,896	53,836	8,189	6
Cornell	183	277	1,703	5
Texas	183	279	1,703	5
Wisconsin	251	450	1,703	5
MAG-US	132,558	697,450	128	20
MAG-CN	101,952	285,561	128	20
MAG-DE	43,032	126,663	128	20
MAG-JP	37,498	90,944	128	20
MAG-RU	32,833	67,994	128	20
MAG-FR	29,262	78,222	128	20

A.2 WEBKB LABEL ALIGNMENT

Table 5: WebKB label mappings before and after correction.

	Before correction			After correction	
Label ID	Cornell	Texas	Wisconsin	Shared mapping	Open-set role
0	course	course	staff	course	Known
1	staff	staff	course	staff	Known
2	faculty	project	student	faculty	Known
3	student	student	faculty	student	Unknown
4	project	faculty	project	project	Unknown

A.3 ADDITIONAL RESULTS

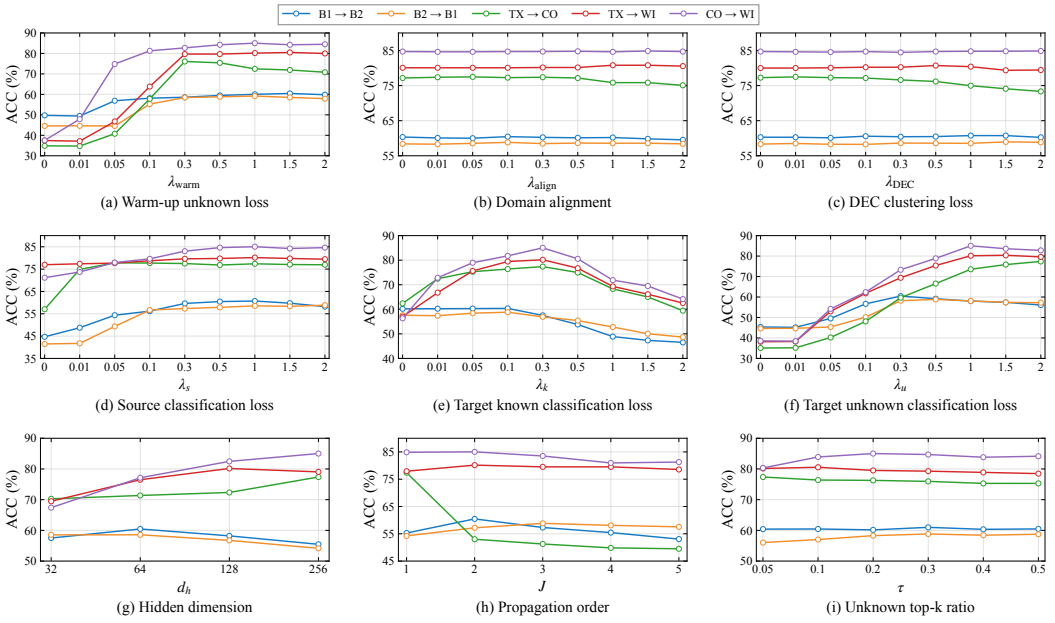
Table 6: Open-set node classification on **Citation**. Results are mean ± standard deviation (%). Bold and underline indicate the best and second-best means.

Task	Metric	Heterophilic GNNs					Closed-set DA						Open-set DA		HODA
		ACM	CMGNN	HeroFilter	H2GCN	HIMP	ADAlign	DiffGDA	Pair-Align	DGSDA	DFT	RSGDA	UAGA	GraphRTA	
C → D	ACC	56.84±0.91	54.47±0.95	53.58±1.34	62.38±0.14	55.90±0.59	68.34±0.90	57.72±3.88	55.68±2.04	63.80±1.49	59.61±1.18	43.77±2.03	52.13±2.85	62.92±0.78	**68.60±2.51**
	HS	60.87±0.65	56.81±1.58	58.00±1.18	66.94±0.17	56.71±1.60	51.13±1.96	58.75±4.05	58.14±1.42	66.36±1.26	61.49±1.64	46.83±1.96	63.72±2.62	63.92±1.01	**70.15±0.99**
	AUC	73.91±0.62	68.71±2.30	67.63±1.67	80.37±0.18	68.55±1.93	73.02±1.34	66.58±4.87	67.41±0.99	78.20±1.58	70.96±1.24	62.40±1.09	77.90±2.35	78.74±0.54	**88.27±0.94**
C → A	ACC	56.35±1.45	53.50±0.60	49.44±0.77	58.20±0.13	49.61±0.48	60.42±0.30	58.49±3.63	54.10±0.98	60.72±1.03	55.98±0.73	43.78±1.48	60.91±0.96	58.75±1.17	**65.64±1.71**
	HS	57.12±1.74	53.61±1.09	51.82±0.64	60.86±0.15	52.37±0.54	56.96±1.14	58.25±5.27	55.97±0.94	60.66±1.77	57.57±1.36	46.23±1.51	58.33±1.05	60.32±1.97	**67.96±1.91**
	AUC	70.24±2.29	61.52±0.71	62.08±1.38	74.90±0.25	65.43±1.37	70.22±0.93	66.67±8.19	66.97±1.55	73.08±2.91	70.77±1.84	61.88±1.36	68.98±2.42	76.93±2.07	**89.08±2.02**
D → C	ACC	62.10±0.57	60.10±1.21	56.45±1.47	62.24±0.13	57.61±1.12	54.97±1.84	56.91±3.97	57.32±1.00	58.06±1.01	60.79±1.17	38.32±6.06	60.41±0.37	58.31±1.08	**70.76±0.92**
	HS	65.13±0.56	62.14±1.94	58.15±1.97	66.00±0.14	60.68±1.31	58.34±1.97	57.85±3.79	58.75±0.48	57.35±1.32	62.72±1.67	40.11±6.71	52.26±0.55	59.98±1.47	**68.93±0.81**
	AUC	77.80±0.85	73.39±1.85	69.01±1.93	78.40±0.46	71.96±1.61	73.24±4.13	62.91±5.26	67.41±0.48	65.30±1.66	75.21±1.52	57.22±8.65	70.48±0.75	71.30±1.91	**90.34±0.65**
D → A	ACC	56.83±0.49	52.49±1.69	48.01±1.63	57.34±0.10	53.71±1.42	55.08±0.70	55.45±4.63	51.82±2.11	55.82±1.22	52.80±0.59	40.14±4.87	60.86±0.52	57.11±1.59	**61.18±1.10**
	HS	59.44±0.58	55.22±1.83	50.06±2.34	60.20±0.10	56.61±1.47	58.17±0.78	53.90±7.30	53.79±2.84	58.12±1.29	55.47±0.73	41.57±5.54	62.13±0.82	59.61±1.37	**64.35±1.16**
	AUC	76.07±0.62	70.18±1.57	63.18±2.89	77.57±0.14	73.75±1.63	77.53±0.37	59.95±11.47	67.59±4.31	74.03±1.97	70.85±1.39	62.09±7.10	71.30±1.43	77.71±1.03	**86.64±1.46**
A → C	ACC	63.71±0.67	63.38±1.21	53.38±1.97	64.53±0.09	55.94±1.07	67.14±1.40	58.28±3.63	58.75±0.58	54.77±0.45	61.68±2.49	40.79±7.37	62.23±0.34	62.25±1.53	**69.22±1.76**
	HS	66.27±0.81	64.41±2.89	54.61±1.33	66.83±0.12	58.40±1.15	67.79±2.88	59.78±4.58	61.39±0.98	55.12±1.56	63.65±5.17	42.44±8.73	59.14±0.33	64.63±1.56	**69.64±0.66**
	AUC	77.04±0.51	72.74±3.69	63.84±1.70	78.92±0.13	68.49±1.41	77.88±3.13	64.54±5.59	69.46±1.17	60.66±2.23	72.55±4.94	59.77±10.86	70.77±0.88	75.43±2.58	**87.79±1.56**
A → D	ACC	61.56±0.45	52.93±0.48	49.73±1.24	59.98±0.12	55.12±0.89	58.81±1.89	58.06±3.63	55.15±1.64	58.69±1.91	54.96±1.24	42.46±3.95	62.23±1.19	59.08±1.83	**62.46±1.13**
	HS	64.97±0.41	57.59±0.50	52.99±1.57	64.30±0.15	58.07±0.98	50.92±6.80	60.51±3.99	58.53±1.75	63.60±1.74	56.82±2.35	43.54±7.66	58.29±1.57	61.93±0.91	**65.33±1.13**
	AUC	78.78±0.33	71.27±0.85	64.68±4.16	80.25±0.05	71.34±1.06	74.09±1.40	68.71±4.25	68.50±1.87	80.56±1.48	65.87±1.92	64.58±8.11	69.89±1.92	78.04±1.16	**83.82±1.16**

Table 7: Open-set node classification on **MAG**. The last class is treated as unknown. Results are mean ± standard deviation (%). Bold and underline indicate the best and second-best means.

Task	Metric	Heterophilic GNNs					Closed-set DA						Open-set DA		HODA
		ACM	CMGNN	HeroFilter	H2GCN	HIMP	ADAlign	DiffGDA	Pair-Align	DGSDA	DFT	RSGDA	UAGA	GraphRTA	
CN → DE	ACC	36.17±0.41	35.06±4.47	33.42±0.41	37.45±0.18	35.16±0.41	35.84±0.35	36.18±0.41	33.44±1.35	34.77±0.43	32.48±0.68	28.24±4.95	34.11±2.51	39.66±1.10	**41.06±0.50**
	HS	34.17±0.32	29.38±3.98	33.06±0.35	35.48±0.18	33.88±0.29	33.06±0.27	31.16±0.27	30.73±0.76	30.54±0.49	30.62±0.65	11.34±1.28	29.19±1.84	33.82±0.70	**38.30±0.35**
	AUC	53.32±0.44	65.54±0.82	50.60±0.38	54.77±0.16	54.99±1.48	60.85±0.32	56.14±0.96	49.28±2.69	59.64±0.25	51.88±2.48	47.48±5.04	62.84±1.33	66.58±0.68	**68.88±0.30**
CN → FR	ACC	35.39±0.17	31.23±0.34	31.47±0.27	35.50±0.09	34.15±0.48	37.55±0.07	33.32±0.70	28.45±1.32	31.70±0.36	32.04±0.23	25.26±4.88	36.16±0.79	38.38±1.90	**40.33±0.18**
	HS	30.31±0.16	29.00±0.06	29.27±0.14	31.56±0.08	30.77±0.40	29.36±0.15	27.77±0.51	28.37±1.43	27.47±0.18	27.41±0.54	11.84±1.49	28.09±1.36	30.12±0.98	**36.49±0.23**
	AUC	54.60±0.13	63.25±0.50	46.99±0.72	58.82±0.06	54.12±1.45	57.07±0.16	60.26±1.00	40.58±2.11	59.50±0.12	47.56±1.05	46.44±5.84	58.36±2.10	65.71±2.02	**70.57±0.50**
CN → JP	ACC	44.65±0.12	44.15±1.64	42.27±0.63	45.74±0.07	41.16±0.66	45.55±0.51	45.03±0.32	47.55±4.49	41.99±0.63	42.72±0.61	38.04±4.58	44.59±1.52	52.40±3.06	**53.70±1.10**
	HS	42.03±0.20	39.57±0.96	40.97±0.38	42.96±0.04	40.76±0.07	41.43±0.26	40.01±0.29	34.56±3.55	39.66±0.37	38.73±0.94	17.86±4.98	37.57±2.03	34.22±5.36	**43.31±0.87**
	AUC	51.50±0.20	58.27±0.52	50.97±0.44	55.64±0.05	51.42±0.72	57.86±0.38	51.78±0.59	50.74±1.10	56.06±0.38	49.44±0.78	47.25±2.86	58.61±0.12	59.62±1.40	**66.55±0.33**
CN → RU	ACC	50.20±0.70	43.65±4.50	43.66±0.61	49.27±0.21	44.73±0.40	46.70±0.21	41.20±1.01	51.97±2.81	45.53±0.70	40.36±1.70	33.42±10.31	55.30±2.53	51.38±7.06	**65.28±1.71**
	HS	37.23±0.52	31.74±0.53	36.85±0.28	38.62±0.19	38.23±0.37	37.83±0.20	34.73±0.32	35.74±0.42	34.84±0.64	35.25±0.31	10.17±1.05	31.78±0.82	33.22±2.33	**39.91±1.86**
	AUC	51.17±0.88	56.89±1.32	47.49±0.47	53.09±0.16	51.51±0.85	55.10±0.17	51.73±0.52	52.43±1.53	52.44±0.19	48.65±0.77	42.32±3.43	64.55±0.64	56.76±2.86	**70.64±0.58**
CN → US	ACC	47.01±0.21	45.45±1.59	40.45±0.67	46.55±0.05	45.96±0.59	46.30±0.26	44.14±0.42	41.69±2.10	44.78±0.13	43.95±0.70	28.74±5.23	44.27±1.53	43.04±2.84	**49.45±0.41**
	HS	45.57±0.07	44.41±1.24	40.42±0.69	46.13±0.04	44.77±0.48	44.60±0.30	42.88±0.24	40.40±1.26	42.18±0.05	42.40±0.75	11.67±2.58	43.28±2.85	39.44±1.19	**48.69±0.62**
	AUC	58.22±0.30	61.63±1.90	50.17±1.21	59.95±0.04	61.85±0.50	61.78±0.59	57.67±0.42	50.45±3.37	60.05±0.09	52.04±1.22	44.95±3.31	64.25±3.67	61.55±1.47	**70.74±0.54**
US → CN	ACC	44.07±0.16	42.57±0.76	44.51±0.82	43.12±0.02	42.25±0.39	41.27±0.34	44.65±0.46	43.02±2.02	37.93±0.25	40.36±3.01	25.42±2.99	40.94±0.52	**48.95±1.29**	47.94±3.27
	HS	38.40±0.15	36.53±1.53	36.02±0.32	39.14±0.01	37.21±0.21	37.73±0.37	37.54±0.30	34.91±1.52	33.53±0.32	33.85±2.46	12.32±3.60	35.24±0.36	27.78±0.95	**39.98±1.99**
	AUC	57.83±0.44	56.44±0.53	50.86±0.55	57.15±0.02	51.88±0.33	57.59±0.23	52.68±0.54	52.53±1.48	50.15±0.57	49.69±2.00	53.76±4.15	49.89±1.17	55.09±1.16	**63.77±1.31**
US → DE	ACC	39.99±0.33	39.49±1.96	36.93±0.37	40.11±0.09	37.91±0.30	40.85±0.16	39.02±0.30	37.53±0.67	38.91±0.08	39.74±0.53	17.29±6.79	41.35±1.44	38.43±1.87	**42.75±0.85**
	HS	38.74±0.29	35.59±2.09	36.24±0.33	38.44±0.05	36.52±0.16	34.95±0.06	33.85±0.18	36.79±0.89	34.22±0.14	35.77±0.67	11.84±4.65	34.20±0.93	36.29±0.59	**41.09±0.40**
	AUC	56.33±0.46	61.18±1.66	54.51±0.69	63.38±0.06	59.86±0.45	64.07±0.28	63.54±0.34	50.48±1.10	64.24±0.17	56.29±1.65	58.69±4.49	68.80±1.40	70.00±0.35	**71.49±0.44**
US → FR	ACC	39.24±0.90	38.49±0.30	37.14±0.46	39.11±0.05	37.14±0.15	36.67±0.48	31.92±0.82	36.98±1.41	32.99±0.24	33.77±0.56	19.60±5.71	41.98±0.83	44.75±1.59	**45.49±0.65**
	HS	34.16±0.39	30.68±0.24	31.84±0.18	34.80±0.03	33.06±0.21	30.23±0.27	28.85±0.52	35.46±1.81	30.20±0.15	32.21±0.43	11.25±2.04	32.40±0.54	25.29±1.83	**38.90±0.41**
	AUC	59.67±0.75	65.88±0.38	57.64±0.45	63.66±0.09	56.34±0.56	61.66±0.33	62.96±0.66	48.85±0.95	63.96±0.51	53.94±1.03	49.96±8.12	70.60±0.84	70.18±0.71	**71.15±1.59**
US → JP	ACC	47.66±0.35	49.55±0.96	44.61±0.51	46.54±0.04	44.03±0.30	44.38±0.70	45.44±0.32	44.03±1.32	46.96±0.22	46.26±0.97	29.76±3.33	49.97±1.77	49.44±3.92	**53.48±0.67**
	HS	43.00±0.26	41.26±0.55	42.15±0.35	44.68±0.04	42.54±0.14	42.68±0.49	41.47±0.30	40.70±1.53	41.24±0.13	41.17±1.46	18.49±2.11	42.24±1.02	37.08±1.98	**45.44±0.78**
	AUC	53.05±0.49	59.83±0.39	52.00±0.55	57.73±0.02	52.04±0.25	60.20±0.68	57.10±0.48	54.15±1.88	57.20±0.19	59.20±1.74	52.98±1.19	63.07±1.42	58.37±1.08	**66.73±0.28**
US → RU	ACC	46.76±0.28	41.54±11.08	46.53±0.58	46.09±0.18	42.72±0.42	49.29±0.34	42.29±1.00	47.50±3.15	45.76±0.62	48.33±1.33	34.46±14.35	52.11±1.54	53.81±5.45	**62.37±1.90**
	HS	38.25±0.10	32.93±8.63	37.17±0.35	37.83±0.17	37.75±0.35	35.27±0.19	35.53±0.23	39.54±1.54	33.88±0.24	37.13±0.54	11.75±2.17	36.84±0.33	36.55±2.27	**40.26±0.99**
	AUC	51.96±0.16	55.67±3.67	49.50±0.26	52.24±0.07	51.17±0.17	56.02±0.12	52.15±0.59	54.19±1.67	54.55±0.39	51.66±0.51	51.98±6.47	65.84±1.35	61.76±1.46	**71.67±0.44**

A.4 Parameter Sensitivity



(a) Warm-up unknown loss (b) Domain alignment (c) DEC clustering loss

(d) Source classification loss (e) Target known classification loss (f) Target unknown classification loss

(g) Hidden dimension (h) Propagation order (i) Unknown top-k ratio

Figure 4: Parameter sensitivity of HODA on Blog and WebKB. Each parameter is varied while the remaining settings are fixed.

A.5 Additional Ablation Study

Table 8 further examines the high-pass views, propagation orders, prototype mixing, and pseudo-label selection. The w/o HP variant removes high-pass views, while 1-hop restricts propagation to the first order. EW-PA replaces reliability-adaptive prototype mixing with an equal weight average of source and target prototypes. Conf. filter replaces consensus with confidence-based selection. Within each classifier-predicted group, nodes are ranked by their predicted-class probability. The

selected nodes provide pseudo-label supervision, and selected unknown nodes are excluded from alignment.

Table 8: Additional ablation study on **Blog and WebKB**. Results are mean ± standard deviation (%) over five seeds. Bold indicates the best mean.

Variant	B1→B2		B2→B1		TX→CO		TX→WI		CO→WI	
	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS
w/o HP	58.36±1.02	57.94±1.11	55.07±0.49	57.79±0.49	74.32±1.25	72.67±1.03	72.03±3.56	68.56±3.62	72.67±1.93	70.57±1.67
1-hop	55.28±0.61	54.09±1.23	54.23±0.26	51.63±0.44	77.38±3.33	76.10±2.89	77.93±3.40	75.45±3.62	84.86±1.78	81.99±1.50
EW-PA	60.02±0.57	61.53±0.51	58.27±0.46	59.75±0.57	76.07±2.88	75.53±2.26	79.04±0.69	75.61±0.95	83.59±2.84	81.05±2.55
Conf. filter	59.92±0.45	61.34±0.42	57.45±0.61	58.85±0.68	74.43±0.64	72.82±0.92	75.62±4.27	73.09±4.90	78.73±1.84	75.80±1.80
HODA	**60.46±0.38**	**61.94±0.39**	**58.83±0.77**	**60.48±0.83**	**77.38±3.33**	**76.10±2.89**	**80.16±2.16**	**77.85±3.56**	**85.02±1.52**	**82.47±1.66**

A.6 PROOFS

A.6.1 MULTISCALE OPEN-SET RISK BOUND

Under the assumptions of the multiscale open-set risk bound, every $h \in \mathcal{H}_J$ satisfies

$$R_t^{\mathrm{os}}(h) \le \min\Big\{(1-\pi_u)\big[R_s(h) + B_J\mathcal{D}_J + 2B_J\kappa_J(\eta+\beta) + \lambda_J\big], \\ L_k(h) + (1-\pi_u)\big[\rho_k\epsilon_{\mathrm{con}} + 1 - \rho_k\big]\Big\} + L_u(h) + \pi_u(1-\rho_u). \tag{14}$$

Moreover, under the nested predictor-class assumption, $\lambda_J \le \lambda_j$ for every $j \le J$.

Proof. Throughout the proof, the encoder and consensus selectors are fixed. For brevity, write $Z = Z^{(J)}$. Define the conditional target risks as

$$R_t^k(h) = \mathbb{E}_t[\ell(h(Z), Y) \mid Y < C], \quad R_t^u(h) = \mathbb{E}_t[\ell(h(Z), C) \mid Y = C].$$

Since all target private labels are mapped to C, the law of total expectation gives

$$R_t^{\mathrm{os}}(h) = (1-\pi_u)R_t^k(h) + \pi_u R_t^u(h). \tag{15}$$

Decompose the target distributions. Let $A = S_u^c$ denote the event that a target node is retained for alignment. Recall that $\eta = \Pr_t(Y = C \mid A)$ is the fraction of unknown nodes among retained target nodes, and $\beta = \Pr_t(S_u \mid Y < C)$ is the fraction of true known-class target nodes excluded from alignment. Define

$$P_{t,k}^{(J)} = P_t(Z \mid Y < C), \quad P_{t,k}^{\mathrm{keep}} = P_t(Z \mid Y < C, A).$$

The second distribution is well defined because

$$\Pr_t(Y < C, A) = \Pr_t(Y < C)\Pr_t(A \mid Y < C) = (1-\pi_u)(1-\beta) > 0.$$

In particular,

$$1 - \eta = \Pr_t(Y < C \mid A) = \frac{(1-\pi_u)(1-\beta)}{\Pr_t(A)} > 0.$$

Thus the retained target distribution contains known-class nodes with positive probability.

If $\eta > 0$, let $P_{t,u}^{\mathrm{keep}} = P_t(Z \mid Y = C, A)$. Conditioning on whether a retained node is known or unknown gives

$$Q_t^{(J)} = (1-\eta)P_{t,k}^{\mathrm{keep}} + \eta P_{t,u}^{\mathrm{keep}}. \tag{16}$$

Similarly, if $\beta > 0$, let $P_{t,k}^{\mathrm{excl}} = P_t(Z \mid Y < C, S_u)$. Conditioning on whether a known-class node is retained or excluded gives

$$P_{t,k}^{(J)} = (1-\beta)P_{t,k}^{\mathrm{keep}} + \beta P_{t,k}^{\mathrm{excl}}. \tag{17}$$

If $\eta = 0$ or $\beta = 0$, the corresponding zero-weight component can be chosen arbitrarily.

Bound the effect of selection on MMD. Let $\mathcal{K}_J$ denote the RKHS associated with k_J, and let $\phi_J(z) = k_J(z,\cdot)$ be its canonical feature map. The reproducing property gives $\|\phi_J(z)\|^2_{\mathcal{K}_J} = k_J(z,z) \le \kappa_J^2$. For any distribution P, write its kernel mean embedding as $m_P = \mathbb{E}_P[\phi_J(Z)]$. The triangle inequality for expectations implies $\|m_P\|_{\mathcal{K}_J} \le \mathbb{E}_P[\|\phi_J(Z)\|_{\mathcal{K}_J}] \le \kappa_J$. Consequently, for any two distributions P and Q,

$$\mathrm{MMD}_{k_J}(P,Q) = \|m_P - m_Q\|_{\mathcal{K}_J} \le \|m_P\|_{\mathcal{K}_J} + \|m_Q\|_{\mathcal{K}_J} \le 2\kappa_J. \tag{18}$$

By linearity of expectation, Eq. (16) implies $m_{Q_t^{(J)}} - m_{P_{t,k}^{\mathrm{keep}}} = \eta\big(m_{P_{t,u}^{\mathrm{keep}}} - m_{P_{t,k}^{\mathrm{keep}}}\big)$. Taking norms and applying Eq. (18) yields

$$\mathrm{MMD}_{k_J}(Q_t^{(J)}, P_{t,k}^{\mathrm{keep}}) = \eta\,\mathrm{MMD}_{k_J}(P_{t,u}^{\mathrm{keep}}, P_{t,k}^{\mathrm{keep}}) \le 2\kappa_J\eta.$$

Likewise, Eq. (17) gives $m_{P_{t,k}^{(J)}} - m_{P_{t,k}^{\mathrm{keep}}} = \beta\big(m_{P_{t,k}^{\mathrm{excl}}} - m_{P_{t,k}^{\mathrm{keep}}}\big)$, and therefore

$$\mathrm{MMD}_{k_J}(P_{t,k}^{(J)}, P_{t,k}^{\mathrm{keep}}) \le 2\kappa_J\beta. \tag{19}$$

Applying the triangle inequality through $Q_t^{(J)}$ and $P_{t,k}^{\mathrm{keep}}$ now gives

$$\begin{aligned}\mathrm{MMD}_{k_J}(P_s^{(J)}, P_{t,k}^{(J)}) &\le \mathrm{MMD}_{k_J}(P_s^{(J)}, Q_t^{(J)}) \\ &\quad + \mathrm{MMD}_{k_J}(Q_t^{(J)}, P_{t,k}^{\mathrm{keep}}) + \mathrm{MMD}_{k_J}(P_{t,k}^{\mathrm{keep}}, P_{t,k}^{(J)}) \\ &\le \mathcal{D}_J + 2\kappa_J(\eta+\beta),\end{aligned} \tag{20}$$

where $\mathcal{D}_J = \mathrm{MMD}_{k_J}(P_s^{(J)}, Q_t^{(J)})$.

Bound known-class risk through alignment. Fix any $h, h' \in \mathcal{H}_J$. For every label pair (z,y),

$$\mathbb{I}(h(z) \ne y) \le \mathbb{I}(h'(z) \ne y) + \mathbb{I}(h(z) \ne h'(z)). \tag{21}$$

Taking expectations over true known-class target nodes yields

$$R_t^k(h) \le R_t^k(h') + \mathbb{E}_{P_{t,k}^{(J)}}[d_{h,h'}(Z)]. \tag{22}$$

By assumption, $d_{h,h'} \in \mathcal{K}_J$ and $\|d_{h,h'}\|_{\mathcal{K}_J} \le B_J$. The reproducing property and the Cauchy-Schwarz inequality therefore give

$$\begin{aligned}\Big|\mathbb{E}_{P_{t,k}^{(J)}}[d_{h,h'}(Z)] - \mathbb{E}_{P_s^{(J)}}[d_{h,h'}(Z)]\Big| &= \Big|\Big\langle d_{h,h'}, m_{P_{t,k}^{(J)}} - m_{P_s^{(J)}}\Big\rangle_{\mathcal{K}_J}\Big| \\ &\le \|d_{h,h'}\|_{\mathcal{K}_J}\|m_{P_{t,k}^{(J)}} - m_{P_s^{(J)}}\|_{\mathcal{K}_J} \le B_J\,\mathrm{MMD}_{k_J}(P_s^{(J)}, P_{t,k}^{(J)}).\end{aligned}$$

Hence

$$\mathbb{E}_{P_{t,k}^{(J)}}[d_{h,h'}(Z)] \le \mathbb{E}_{P_s^{(J)}}[d_{h,h'}(Z)] + B_J\,\mathrm{MMD}_{k_J}(P_s^{(J)}, P_{t,k}^{(J)}).$$

On a source example, disagreement implies that at least one predictor is incorrect: $d_{h,h'}(Z) \le \ell(h(Z),Y) + \ell(h'(Z),Y)$. Taking source expectations gives $\mathbb{E}_{P_s^{(J)}}[d_{h,h'}(Z)] \le R_s(h) + R_s(h')$. Substituting this inequality and Eq. (20) into Eq. (22) yields

$$R_t^k(h) \le R_s(h) + B_J\mathcal{D}_J + 2B_J\kappa_J(\eta+\beta) + R_s(h') + R_t^k(h').$$

For every $\varepsilon > 0$, the definition of λ_J provides a predictor $h' \in \mathcal{H}_J$ such that $R_s(h') + R_t^k(h') \le \lambda_J + \varepsilon$. Using this predictor and letting $\varepsilon \to 0$ gives

$$R_t^k(h) \le R_s(h) + B_J\mathcal{D}_J + 2B_J\kappa_J(\eta+\beta) + \lambda_J. \tag{23}$$

Bound known-class risk through consensus. The known-class contribution to target risk can be split according to S_k:

$$\begin{aligned}(1-\pi_u)R_t^k(h) &= \mathbb{E}_t[\mathbb{I}(Y<C)\mathbb{I}(S_k)\ell(h(Z),Y)] \\ &\quad + \mathbb{E}_t[\mathbb{I}(Y<C)\mathbb{I}(S_k^c)\ell(h(Z),Y)].\end{aligned} \tag{24}$$

On S_k, the agreed label $\widetilde{Y}$ satisfies $\ell(h(Z), Y) \le \ell(h(Z), \widetilde{Y}) + \mathbb{I}(\widetilde{Y} \neq Y)$. On S_k^c, the zero-one loss is at most one. Applying these two bounds to Eq. (24) gives

$$(1-\pi_u)R_t^k(h) \le \mathbb{E}_t[\mathbb{I}(Y<C)\mathbb{I}(S_k)\ell(h(Z),\widetilde{Y})] + \Pr_t(Y<C, S_k, \widetilde{Y}\neq Y) + \Pr_t(Y<C, S_k^c). \tag{25}$$

The first term is bounded by $L_k(h)$:

$$\mathbb{E}_t[\mathbb{I}(Y<C)\mathbb{I}(S_k)\ell(h(Z),\widetilde{Y})] \le \mathbb{E}_t[\mathbb{I}(S_k)\ell(h(Z),\widetilde{Y})] = L_k(h).$$

If $\Pr_t(S_k, Y<C) > 0$, the second term is

$$\begin{aligned}\Pr_t(Y<C, S_k, \widetilde{Y}\neq Y) &= \Pr_t(Y<C)\Pr_t(S_k \mid Y<C) \times \Pr_t(\widetilde{Y}\neq Y \mid S_k, Y<C)\\ &= (1-\pi_u)\rho_k\epsilon_{\mathrm{con}}.\end{aligned}$$

If $\Pr_t(S_k, Y<C) = 0$, both sides are zero by the convention defining ϵ_{con}.

The third term is $\Pr_t(Y<C, S_k^c) = (1-\pi_u)\Pr_t(S_k^c \mid Y<C) = (1-\pi_u)(1-\rho_k)$. Substituting these expressions into Eq. (25) yields

$$(1-\pi_u)R_t^k(h) \le L_k(h) + (1-\pi_u)\big[\rho_k\epsilon_{\mathrm{con}} + 1 - \rho_k\big]. \tag{26}$$

Bound the unknown-class contribution. Since every unknown label equals C,

$$\begin{aligned}\pi_u R_t^u(h) &= \mathbb{E}_t[\mathbb{I}(Y=C)\ell(h(Z),C)]\\ &= \mathbb{E}_t[\mathbb{I}(Y=C)\mathbb{I}(S_u)\ell(h(Z),C)] + \mathbb{E}_t[\mathbb{I}(Y=C)\mathbb{I}(S_u^c)\ell(h(Z),C)].\end{aligned}$$

The selected-node term is bounded by

$$\mathbb{E}_t[\mathbb{I}(Y=C)\mathbb{I}(S_u)\ell(h(Z),C)] \le \mathbb{E}_t[\mathbb{I}(S_u)\ell(h(Z),C)] = L_u(h).$$

For the unselected-node term, boundedness of the zero-one loss gives

$$\mathbb{E}_t[\mathbb{I}(Y=C)\mathbb{I}(S_u^c)\ell(h(Z),C)] \le \Pr_t(Y=C, S_u^c) = \pi_u \Pr_t(S_u^c \mid Y=C) = \pi_u(1-\rho_u).$$

Therefore,

$$\pi_u R_t^u(h) \le L_u(h) + \pi_u(1-\rho_u). \tag{27}$$

Combine the risk bounds. Multiplying Eq. (23) by $1-\pi_u$ gives one upper bound on $(1-\pi_u)R_t^k(h)$. Equation (26) gives another upper bound on the same quantity. Since both hold simultaneously,

$$(1-\pi_u)R_t^k(h) \le \min\Big\{(1-\pi_u)\big[R_s(h) + B_J\mathcal{D}_J + 2B_J\kappa_J(\eta+\beta) + \lambda_J\big], \; L_k(h) + (1-\pi_u)\big[\rho_k\epsilon_{\mathrm{con}} + 1 - \rho_k\big]\Big\}. \tag{28}$$

Adding Eq. (27) and using Eq. (15) proves the stated open-set risk bound.

Compare the ideal risks across scales. Fix $j \le J$, and let Π_j project $Z^{(J)}$ onto the views in $Z^{(j)}$. For every $g \in \mathcal{H}_j$, the nested predictor-class assumption ensures that $\bar{g} = g \circ \Pi_j$ belongs to $\mathcal{H}_J$. Since $\Pi_j(Z^{(J)}) = Z^{(j)}$, the two predictors make identical predictions: $\bar{g}(Z^{(J)}) = g(Z^{(j)})$. Their source and known-class target risks therefore coincide. Because λ_J is the infimum over all predictors in $\mathcal{H}_J$,

$$\begin{aligned}\lambda_J &\le \mathbb{E}_s[\ell(\bar{g}(Z^{(J)}), Y)] + \mathbb{E}_t[\ell(\bar{g}(Z^{(J)}), Y) \mid Y<C]\\ &= \mathbb{E}_s[\ell(g(Z^{(j)}), Y)] + \mathbb{E}_t[\ell(g(Z^{(j)}), Y) \mid Y<C].\end{aligned}$$

This inequality holds for every $g \in \mathcal{H}_j$. Taking the infimum over g gives $\lambda_J \le \lambda_j$, completing the proof.

A.6.2 Multi-Hop High-Pass Filtering under Heterophily

Under the assumptions of Theorem 2, for every $j \geq 1$,

$$\|\mathbf{P}_{a,H}^{j}\mathbf{v}_a\|_2 = (1-\vartheta_a)^j\|\mathbf{v}_a\|_2 > \vartheta_a^j\|\mathbf{v}_a\|_2 = \|\mathbf{P}_{a,L}^{j}\mathbf{v}_a\|_2.$$

If the interference signal satisfies the stated spectral separation condition, then

$$\frac{\|\mathbf{P}_{a,H}^{j}\mathbf{e}_a\|_2}{\|\mathbf{P}_{a,H}^{j}\mathbf{v}_a\|_2} \leq \left(\frac{1-\nu_{a,0}}{2(1-\vartheta_a)}\right)^j \frac{\|\mathbf{e}_a\|_2}{\|\mathbf{v}_a\|_2}.$$

Proof. Recall that $\mathbf{v}_a \in \{-1,1\}^{n_a}$ encodes the two classes, and $\vartheta_a < 1/2$ is the common fraction of same-class neighbors. Let $d_a > 0$ be the common node degree. Since the graph is regular, $\mathbf{D}_a = d_a\mathbf{I}_{n_a}$ and $\tilde{\mathbf{A}}_a = \mathbf{A}_a/d_a$.

For each node i, same-class neighbors have signal value $v_{a,i}$, while different-class neighbors have value $-v_{a,i}$. Therefore,

$$(\tilde{\mathbf{A}}_a\mathbf{v}_a)_i = \frac{1}{d_a}\sum_{m\in\mathcal{N}(i)} v_{a,m} = \frac{d_a\vartheta_a v_{a,i} - d_a(1-\vartheta_a)v_{a,i}}{d_a} = (2\vartheta_a - 1)v_{a,i}.$$

Thus $\tilde{\mathbf{A}}_a\mathbf{v}_a = (2\vartheta_a-1)\mathbf{v}_a$. Using the definitions of the filters gives

$$\mathbf{P}_{a,L}\mathbf{v}_a = \frac{1}{2}(\mathbf{v}_a + \tilde{\mathbf{A}}_a\mathbf{v}_a) = \vartheta_a\mathbf{v}_a, \quad \mathbf{P}_{a,H}\mathbf{v}_a = \frac{1}{2}(\mathbf{v}_a - \tilde{\mathbf{A}}_a\mathbf{v}_a) = (1-\vartheta_a)\mathbf{v}_a.$$

Repeated application therefore yields

$$\mathbf{P}_{a,L}^{j}\mathbf{v}_a = \vartheta_a^j\mathbf{v}_a, \quad \mathbf{P}_{a,H}^{j}\mathbf{v}_a = (1-\vartheta_a)^j\mathbf{v}_a.$$

Since $0 \leq \vartheta_a < 1/2$ and $\|\mathbf{v}_a\|_2 = \sqrt{n_a} > 0$, taking norms gives

$$\|\mathbf{P}_{a,H}^{j}\mathbf{v}_a\|_2 = (1-\vartheta_a)^j\|\mathbf{v}_a\|_2 > \vartheta_a^j\|\mathbf{v}_a\|_2 = \|\mathbf{P}_{a,L}^{j}\mathbf{v}_a\|_2.$$

Next, consider the interference signal $\mathbf{e}_a$. Because $\tilde{\mathbf{A}}_a$ is symmetric, it has an orthonormal eigenbasis $\{\mathbf{u}_{a,m}\}_{m=1}^{n_a}$ with eigenvalues $\nu_{a,m} \in [-1,1]$. By assumption, $\mathbf{e}_a$ is supported only on eigenvectors with eigenvalues at least $\nu_{a,0}$, where $2\vartheta_a - 1 < \nu_{a,0} \leq 1$. Hence

$$\mathbf{e}_a = \sum_{m:\nu_{a,m}\geq\nu_{a,0}} b_{a,m}\mathbf{u}_{a,m}, \quad b_{a,m} = \mathbf{u}_{a,m}^{\top}\mathbf{e}_a.$$

For each eigenvector, $\mathbf{P}_{a,H}\mathbf{u}_{a,m} = \frac{1-\nu_{a,m}}{2}\mathbf{u}_{a,m}$. Consequently,

$$\mathbf{P}_{a,H}^{j}\mathbf{e}_a = \sum_{m:\nu_{a,m}\geq\nu_{a,0}} b_{a,m}\left(\frac{1-\nu_{a,m}}{2}\right)^j \mathbf{u}_{a,m}.$$

Orthonormality implies

$$\|\mathbf{P}_{a,H}^{j}\mathbf{e}_a\|_2^2 = \sum_{m:\nu_{a,m}\geq\nu_{a,0}} b_{a,m}^2\left(\frac{1-\nu_{a,m}}{2}\right)^{2j} \leq \left(\frac{1-\nu_{a,0}}{2}\right)^{2j}\sum_{m:\nu_{a,m}\geq\nu_{a,0}} b_{a,m}^2 = \left(\frac{1-\nu_{a,0}}{2}\right)^{2j}\|\mathbf{e}_a\|_2^2.$$

The inequality uses $0 \leq 1-\nu_{a,m} \leq 1-\nu_{a,0}$ on the interference support. Taking square roots gives

$$\|\mathbf{P}_{a,H}^{j}\mathbf{e}_a\|_2 \leq \left(\frac{1-\nu_{a,0}}{2}\right)^j\|\mathbf{e}_a\|_2.$$

Since $\|\mathbf{P}_{a,H}^{j}\mathbf{v}_a\|_2 = (1-\vartheta_a)^j\|\mathbf{v}_a\|_2 > 0$, division yields

$$\frac{\|\mathbf{P}_{a,H}^{j}\mathbf{e}_a\|_2}{\|\mathbf{P}_{a,H}^{j}\mathbf{v}_a\|_2} \leq \left(\frac{1-\nu_{a,0}}{2(1-\vartheta_a)}\right)^j \frac{\|\mathbf{e}_a\|_2}{\|\mathbf{v}_a\|_2}.$$

Finally, $2\vartheta_a - 1 < \nu_{a,0} \leq 1$ implies $0 \leq 1-\nu_{a,0} < 2(1-\vartheta_a)$, so the factor $(1-\nu_{a,0})/[2(1-\vartheta_a)]$ lies in $[0,1)$. The upper bound therefore converges to zero as j increases.

A.6.3 SOURCE-CONSTRAINED PROTOTYPE ADAPTATION

Under the assumptions of Theorem 3, the unique minimizer of $\|\boldsymbol{\mu} - \boldsymbol{\mu}_c^s\|_2^2 + r_c\|\boldsymbol{\mu} - \boldsymbol{\mu}_c^t\|_2^2$ is $\boldsymbol{\mu}_c = (\boldsymbol{\mu}_c^s + r_c\boldsymbol{\mu}_c^t)/(1 + r_c)$. It satisfies

$$\|\boldsymbol{\mu}_c - \boldsymbol{m}_{t,c}\|_2 \leq (b_c + r_c e_c)/(1 + r_c).$$

Proof. Fix a known class c with $\sum_i w_{i,c} > 0$, and treat the prototypes and their reliability score as constants. Recall that $w_{i,c} = p_{t,i,c}q_{t,i,c}^s$ and $r_c = (\sum_i w_{i,c})/(\sum_i p_{t,i,c})$. Since $0 \leq q_{t,i,c}^s \leq 1$,

$$0 < \sum_i w_{i,c} \leq \sum_i p_{t,i,c}.$$

Thus the denominator is positive and $0 < r_c \leq 1$. Define $f_c(\boldsymbol{\mu}) = \|\boldsymbol{\mu} - \boldsymbol{\mu}_c^s\|_2^2 + r_c\|\boldsymbol{\mu} - \boldsymbol{\mu}_c^t\|_2^2$. Differentiating with respect to $\boldsymbol{\mu}$ gives $\nabla f_c(\boldsymbol{\mu}) = 2(\boldsymbol{\mu} - \boldsymbol{\mu}_c^s) + 2r_c(\boldsymbol{\mu} - \boldsymbol{\mu}_c^t)$. The stationary-point condition is therefore $(1 + r_c)\boldsymbol{\mu} = \boldsymbol{\mu}_c^s + r_c\boldsymbol{\mu}_c^t$. Because $1 + r_c > 0$, its solution is

$$\boldsymbol{\mu}_c = (\boldsymbol{\mu}_c^s + r_c\boldsymbol{\mu}_c^t)/(1 + r_c).$$

Moreover, $\nabla^2 f_c(\boldsymbol{\mu}) = 2(1 + r_c)\mathbf{I}$. For every nonzero vector $\mathbf{u}$,

$$\mathbf{u}^\top \nabla^2 f_c(\boldsymbol{\mu})\mathbf{u} = 2(1 + r_c)\|\mathbf{u}\|_2^2 > 0.$$

Hence f_c is strictly convex, and the stationary point is its unique global minimizer.

Recall that $\boldsymbol{m}_{t,c} = \mathbb{E}_t[\mathbf{z} \mid Y = c]$ is the true target class center, assumed to exist. The source and target prototype errors are $b_c = \|\boldsymbol{\mu}_c^s - \boldsymbol{m}_{t,c}\|_2$ and $e_c = \|\boldsymbol{\mu}_c^t - \boldsymbol{m}_{t,c}\|_2$. Subtracting $\boldsymbol{m}_{t,c}$ from the update gives

$$\boldsymbol{\mu}_c - \boldsymbol{m}_{t,c} = \frac{\boldsymbol{\mu}_c^s + r_c\boldsymbol{\mu}_c^t - (1 + r_c)\boldsymbol{m}_{t,c}}{1 + r_c} = \frac{(\boldsymbol{\mu}_c^s - \boldsymbol{m}_{t,c}) + r_c(\boldsymbol{\mu}_c^t - \boldsymbol{m}_{t,c})}{1 + r_c}.$$

Since $r_c > 0$, the triangle inequality yields

$$\|\boldsymbol{\mu}_c - \boldsymbol{m}_{t,c}\|_2 \leq \frac{\|\boldsymbol{\mu}_c^s - \boldsymbol{m}_{t,c}\|_2 + r_c\|\boldsymbol{\mu}_c^t - \boldsymbol{m}_{t,c}\|_2}{1 + r_c} = \frac{b_c + r_c e_c}{1 + r_c}.$$

If $e_c < b_c$, then

$$\frac{b_c + r_c e_c}{1 + r_c} = b_c - \frac{r_c(b_c - e_c)}{1 + r_c} < b_c.$$

Thus the adapted prototype is strictly closer to the true target class center than the source prototype whenever the target estimate is more accurate.

A.7 ALGORITHM

Algorithm 1 HODA

Input: Source graph $\mathcal{G}_s$; target graph $\mathcal{G}_t$; propagation order J; epochs T_w, T_a; candidate ratio $\tau \in (0, 1]$; prototype update interval Δ and rate $\gamma \in (0, 1]$; loss weights.
Output: Target predictions $\widehat{\mathbf{y}}_t$.
Initialize the shared encoder and $(C + 1)$-class classifier.
Construct identity, low-pass, and high-pass views for both domains up to order J.
for $e = 1, \ldots, T_w$ **do** ▷ Source warm-up
    Compute source representations and classifier probabilities $\mathbf{P}_s$.
    Compute $\mathcal{L}_{\mathrm{src}}$ and $\mathcal{L}_{\mathrm{leaveout}}$.
    Update the encoder and classifier using $\mathcal{L}_{\mathrm{warm}}$.
end for
Compute $\mathbf{Z}_t, \mathbf{P}_t$ in evaluation mode.
Set $k = \lceil \tau n_t \rceil$ and select the top k nodes by $p_{t,i,C}$ as $\mathcal{U}$.
Initialize $\boldsymbol{\mu}_C = \mathrm{mean}_{i \in \mathcal{U}}\, \mathbf{z}_{t,i}$. ▷ Unknown prototype
for $e = 1, \ldots, T_a$ **do** ▷ Open-set adaptation
    Compute $\mathbf{Z}_s, \mathbf{Z}_t, \mathbf{P}_s, \mathbf{P}_t$, and channel representations $\mathbf{H}_{a,r}$.
    Compute source prototypes $\boldsymbol{\mu}_c^s$ and assignments $q_{t,i,c}^s$.
    Compute weights $w_{i,c} = p_{t,i,c} q_{t,i,c}^s$.
    for $c = 0, \ldots, C - 1$ **do**
        if $\sum_i w_{i,c} > 0$ **then**
            Compute $\boldsymbol{\mu}_c^t = (\sum_i w_{i,c} \mathbf{z}_{t,i}) / (\sum_i w_{i,c})$ and $r_c = (\sum_i w_{i,c}) / (\sum_i p_{t,i,c})$.
            Set $\boldsymbol{\mu}_c = (\boldsymbol{\mu}_c^s + r_c \boldsymbol{\mu}_c^t)/(1 + r_c)$. ▷ Prototype adaptation
        else
            Set $\boldsymbol{\mu}_c = \boldsymbol{\mu}_c^s$.
        end if
    end for
    Detach the known-class prototypes.
    Compute assignments $q_{t,i,c}$ and detached sharpened targets $\widehat{q}_{t,i,c}$.
    Compute $\mathcal{L}_{\mathrm{DEC}}$.
    Select matching classifier and clustering predictions. ▷ Consensus selection
    Compute $\mathcal{L}_{\mathrm{con}}$, setting empty-set terms to zero.
    Exclude jointly predicted unknown nodes and compute $\mathcal{L}_{\mathrm{align}}$ across the three channels.
    Compute current $\mathcal{L}_{\mathrm{src}}$ and the total objective $\mathcal{L}$.
    Update the encoder, classifier, and $\boldsymbol{\mu}_C$ using Adam.
    if $e \bmod \Delta = 0$ and $e < T_a$ **then**
        Recompute $\mathbf{Z}_t, \mathbf{P}_t$ in evaluation mode and refresh $\mathcal{U}$.
        Without gradient tracking, set $\boldsymbol{\mu}_C \leftarrow (1 - \gamma)\boldsymbol{\mu}_C + \gamma\, \mathrm{mean}_{i \in \mathcal{U}}\, \mathbf{z}_{t,i}$.
    end if
end for
Compute final $\mathbf{P}_t$ in evaluation mode.
return $\widehat{y}_{t,i} = \arg\max_{c \in \{0,\ldots,C\}} p_{t,i,c}$ for all target nodes.