OPEN-SET CROSS-NETWORK NODE CLASSIFICATION VIA OOD-GUIDED SELECTIVE ALIGNMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-network node classification transfers knowledge from a labeled source graph to an unlabeled target graph under distribution shifts. In open-set settings, aligning target nodes from unseen classes with known source classes can cause negative transfer. Existing open-set approaches primarily rely on prediction-related signals to guide adaptation, leaving distributional evidence insufficiently exploited for unknown identification. We propose OGSA, an out-of-distribution (OOD)-guided selective alignment framework that uses distributional deviation as complementary evidence for identifying unknown target nodes. Specifically, agreement between classifier predictions and prototype-initialized clustering provides initial target pseudo-labels. For nodes with conflicting predictions, structure-aware OOD scores provide additional evidence to identify potential unknown nodes. The refined assignments supervise target predictions and guide signed domain adversarial learning. Experiments on diverse graph benchmarks show that OGSA achieves competitive performance in both known class classification and unknown class recognition under open-set domain shifts.

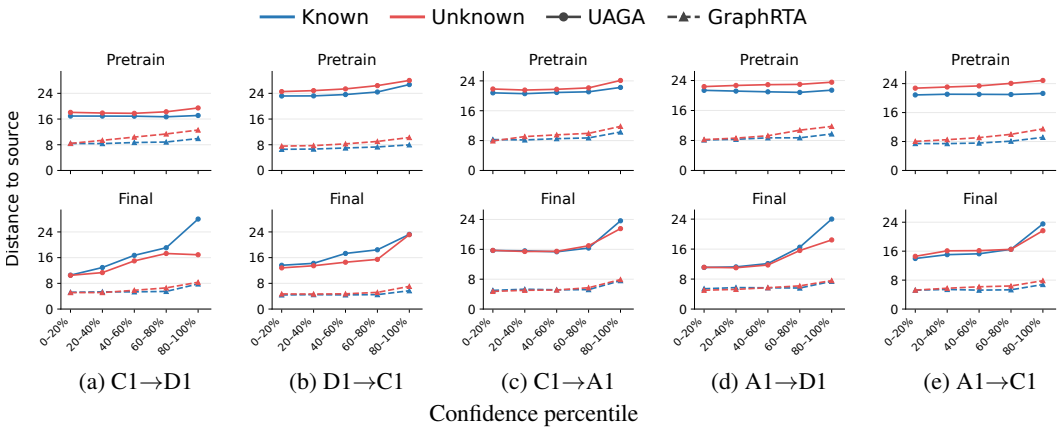
1 INTRODUCTION

Graph neural networks (GNNs) combine node attributes and graph structure to learn node representations (Kipf & Welling, 2016; Hamilton et al., 2017; Veličković et al., 2017). Since labeling each new graph is costly, transferring knowledge from a labeled source graph to an unlabeled target graph is desirable. However, feature and structural differences limit generalization across graphs, motivating graph domain adaptation (GDA) (Shen et al., 2020b; Wu et al., 2020a; Yang et al., 2022). Many GDA methods assume a shared label space, whereas target graphs may contain classes absent from the source. In this open-set setting (Wang et al., 2024; Shen et al., 2025), aligning unknown nodes with source classes can cause negative transfer (Saito et al., 2018; Wang et al., 2019).

Existing open-set methods address this challenge by identifying potentially unknown target nodes to guide adaptation. UAGA (Shen et al., 2025) combines unknown class probabilities with clustering initialized by source prototypes, using the inferred pseudo-labels to control the direction of domain adversarial learning. GraphRTA (Zhang & He, 2025) estimates known and unknown weights through a mixture model of prediction entropy to guide model and graph reprogramming. Both primarily rely on prediction signals to handle distribution discrepancies, overlooking their potential as additional evidence for unknown node identification. Figure 1 shows that, after pretraining, unknown target nodes generally have larger Mahalanobis distances (Lee et al., 2018) to source known class distributions than known nodes within the same confidence quantile intervals. The gap is even larger in higher confidence groups, highlighting its value for distinguishing known from unknown nodes even at high prediction confidence. This contrast weakens or reverses after full adaptation, suggesting that the signal remains insufficiently exploited during training. These observations motivate incorporating distribution discrepancies into unknown node identification.

This perspective relates to out-of-distribution (OOD) detection, which identifies unfamiliar inputs through deviation from a reference distribution (Lee et al., 2018; Fuchsgruber et al., 2024). However, deviation may reflect either a domain shift within shared classes or the presence of unknown classes in our setting. The former calls for alignment, while the latter requires avoiding inappropriate alignment. OOD evidence alone is therefore insufficient. We combine it with class predictions and graph structure to identify potentially unknown nodes and selectively align transferable ones.

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.



(a) C1→D1 (b) D1→C1 (c) C1→A1 (d) A1→D1 (e) A1→C1

Figure 1: Target nodes are divided into five confidence quantile groups. Confidence is the maximum softmax probability over the known classes. Within each group, we compute the mean Mahalanobis distance to the source known class distributions separately for known and unknown target nodes. Pretrain denotes 30 epochs of training with labeled source nodes and an adversarial binary cross-entropy objective for unlabeled target nodes, following Shen et al. (2025). Final denotes representations after full adaptation.

Unlike a common OOD detection pipeline applied after training (Lee et al., 2018), we need unknown node estimates from the outset of adaptation to determine which nodes to align. Since source labels provide no direct supervision for unknown classes, auxiliary training signals are needed before selective alignment begins.

To this end, we propose OGSA, an OOD-guided selective alignment framework for open-set cross-network node classification. We initialize unknown class prediction through source supervision and auxiliary objectives on both domains. A source regularizer masks each node's true class logit and encourages the unknown output over the remaining incorrect classes, while a target adversarial objective supplies an unknown prediction signal without target labels (Saito et al., 2018; Shen et al., 2025). During adaptation, we combine class predictions and prototype-initialized clustering with structure-aware OOD estimation. The resulting unknown node assignments provide additional supervision and guide domain adversarial learning through signed gradient reversal (Shen et al., 2025). Nodes inferred as known are aligned with the source domain, while those inferred as unknown are pushed away from it.

Our contributions are summarized as follows:

- We identify source-distribution compatibility as a complementary signal for unknown-node identification in open-set graph domain adaptation. We show that known and unknown target nodes exhibit systematically different deviations from source known-class distributions even at comparable prediction confidence, revealing information that prediction-based signals alone fail to capture.
- We propose OGSA, an OOD-guided selective alignment framework that combines classifier predictions, prototype-initialized clustering, and structure-aware OOD evidence to refine ambiguous target assignments. The refined assignments are then used for both target supervision and signed domain adversarial learning, aligning transferable target nodes while repelling likely unknown nodes.
- We conduct extensive experiments across 15 cross-network transfer tasks and show that incorporating OOD evidence consistently improves both known-class classification and unknown-class recognition. Ablation and sensitivity studies further demonstrate that the gains come from using distributional evidence as a complement to prediction-based signals.

2 Related Work

Graph Domain Adaptation. Graph neural networks (GNNs) learn node representations by aggregating neighborhood information (Kipf & Welling, 2016; Hamilton et al., 2017; Veličković et al., 2017; Wu et al., 2019; Li et al., 2019; Chen et al., 2020), but feature and structural shifts limit their generalization across graphs. Graph domain adaptation addresses this issue by transferring knowledge from a labeled source graph to an unlabeled target graph. Existing approaches employ adversarial alignment (Shen et al., 2020a), joint modeling of local and global information with distribution alignment (Wu et al., 2020a), and strategies that account for structural differences (Wu et al., 2022; Yang et al., 2022; Liu et al., 2024b). Common alignment techniques include domain adversarial learning (Ganin et al., 2016) and maximum mean discrepancy (Gretton et al., 2012). Many methods assume a shared label space, making them vulnerable to negative transfer when unknown target classes are aligned with source classes.

Open-set Graph Domain Adaptation. Open-set node classification recognizes known classes while detecting unseen ones. OpenWGL (Wu et al., 2020b) learns uncertain representations through a variational graph model, while G2Pxy (Zhang et al., 2023) generates proxy unknown nodes through mixup. Neither directly addresses adaptation across graphs. In broader domain adaptation, DANCE (Saito et al., 2020) combines neighborhood clustering with entropy separation. For graph transfer, UDANE (Chen et al., 2023) uses contrastive learning within and across domains, and SDA (Wang et al., 2024) employs group-wise alignment. UAGA (Shen et al., 2025) combines unknown class probabilities with prototype-initialized clustering to guide adversarial alignment. GraphRTA (Zhang & He, 2025) uses prediction entropy to weight known and unknown nodes during model and graph reprogramming. Both primarily use prediction-related signals to handle distribution discrepancies, overlooking source distribution compatibility as additional evidence for unknown identification and alignment decisions.

Out-of-Distribution Detection. OOD detection identifies inputs that differ from the training distribution using softmax confidence (Hendrycks & Gimpel, 2016), temperature scaling and input perturbations (Liang et al., 2017), energy scores (Liu et al., 2020), or distances to class distributions (Lee et al., 2018). Graph OOD methods further exploit node representations and graph structure to estimate uncertainty and distributional deviation (Hart et al., 2023; Ma et al., 2024; Fuchsgruber et al., 2024). These methods assume that in-distribution nodes follow the training distribution, whereas shared-class target nodes also deviate from the source domain in our setting. OGSA therefore uses OOD evidence together with class predictions rather than as a standalone detector.

3 Theoretical Analysis

3.1 Preliminaries

For each domain $r \in \{s, t\}$, where s and t denote the source and target domains, let $\mathcal{G}_r = (\mathbf{A}_r, \mathbf{X}_r)$ denote a graph with node set V_r, adjacency matrix $\mathbf{A}_r \in \mathbb{R}^{|V_r|\times|V_r|}$, and node features $\mathbf{X}_r \in \mathbb{R}^{|V_r|\times d}$. The source graph is labeled with $\mathbf{Y}_s = (y_{s,i})_{i=1}^{|V_s|}$ over the label space $\mathcal{Y}_s = \{0, \dots, K-1\}$, while the target graph is unlabeled with true label space $\mathcal{Y}_t \supsetneq \mathcal{Y}_s$. The unknown classes $\mathcal{Y}_t \setminus \mathcal{Y}_s$ are collectively assigned output label K. For a node with true label Y, define $\widetilde{Y} = Y$ if $Y \in \mathcal{Y}_s$ and $\widetilde{Y} = K$ otherwise. The objective is to correctly classify target nodes from shared classes and assign nodes from unknown classes to label K. The framework contains a graph encoder g_θ, a classifier f_ψ, and a domain discriminator D_ϕ. The shared encoder produces $\mathbf{H}_r = g_\theta(\mathbf{A}_r, \mathbf{X}_r) \in \mathbb{R}^{|V_r|\times d_h}$ with representation dimension d_h, and the classifier produces logits $\mathbf{Z}_r = f_\psi(\mathbf{H}_r) \in \mathbb{R}^{|V_r|\times(K+1)}$, whose i-th rows are denoted by $\mathbf{h}_{r,i}$ and $\mathbf{z}_{r,i}$. Let $F(\mathbf{h}) = \arg\max_{0\le c\le K}[f_\psi(\mathbf{h})]_c$ denote the model prediction and $F^{\mathrm{kn}}(\mathbf{h}) = \arg\max_{0\le c<K}[f_\psi(\mathbf{h})]_c$ its prediction restricted to known classes, abbreviated as $F(i)$ and $F^{\mathrm{kn}}(i)$ for node i.

3.2 Target Risk Bound

We analyze a fixed adaptation stage using the loss $\ell(a, b) = \mathbf{1}[a \neq b]$. Probabilities and expectations indexed by s and t refer to uniform node sampling from the corresponding graphs. Let $(\mathbf{h}, Y)$ denote

the representation and true label of a sampled node. The target unknown proportion is $\pi_{\mathrm{unk}} = \Pr_t(\widetilde{Y} = K) \in (0, 1)$.

Definition 1 (Classification risks and domain discrepancy). *Let $\mathcal{H}$ be a hypothesis class of predictors from the fixed node representations to $\mathcal{Y}_s$, containing F^{kn}. For $q \in \mathcal{H}$, the source classification risk and target known class risk are $R_s(q) = \mathbb{E}_s \ell(q(\mathbf{h}), Y)$ and $R_{t,\mathrm{kn}}(q) = \mathbb{E}_t[\ell(q(\mathbf{h}), Y) \mid \widetilde{Y} < K]$. The open-set target risk is $R_t^{\mathrm{os}}(F) = \mathbb{E}_t \ell(F(\mathbf{h}), \widetilde{Y})$. Define the minimum joint known class risk as $\Lambda_{\mathrm{kn}} = \inf_{q \in \mathcal{H}} \{R_s(q) + R_{t,\mathrm{kn}}(q)\}$ and the shared-class domain discrepancy as*

$$\mathcal{D}_{\mathcal{H}} = \sup_{q, q' \in \mathcal{H}} \left| \mathbb{E}_s \ell(q(\mathbf{h}), q'(\mathbf{h})) - \mathbb{E}_{t,\mathrm{kn}} \ell(q(\mathbf{h}), q'(\mathbf{h})) \right|,$$

where $\mathbb{E}_{t,\mathrm{kn}}$ conditions on $\widetilde{Y} < K$. This discrepancy measures the largest cross-domain difference in predictor disagreement.

Definition 2 (Unknown selection and OOD refinement). *Let $b_i^0 \in \{0, 1\}$ indicate whether prediction and clustering initially assign target node i to the unknown class. Thus, $b_i^0 = 0$ includes both known assignments and unresolved nodes. Let $\mathcal{C} \subseteq \{i : b_i^0 = 0\}$ contain unresolved candidates eligible for refinement. For an OOD score E_i with larger values indicating stronger distributional deviation and a threshold $\tau \in \mathbb{R}$, define the additional unknown set $\Omega_\tau = \{i \in \mathcal{C} : E_i \geq \tau\}$ and the refined indicator $b_i^\tau = b_i^0 \vee \mathbf{1}[i \in \Omega_\tau]$. The auxiliary predictor $F_\tau(i)$ returns K when $b_i^\tau = 1$ and $F^{\mathrm{kn}}(i)$ otherwise. The initial known rejection and unknown acceptance rates are $\alpha_0 = \Pr_t(b_i^0 = 1 \mid \widetilde{Y} < K)$ and $\beta_0 = \Pr_t(b_i^0 = 0 \mid \widetilde{Y} = K)$. The additional known rejection mass $m_{\mathrm{kn}}(\tau) = \Pr_t(i \in \Omega_\tau, \widetilde{Y} < K)$ and unknown recovery mass $m_{\mathrm{unk}}(\tau) = \Pr_t(i \in \Omega_\tau, \widetilde{Y} = K)$ are the fractions of all target nodes that are newly assigned as unknown but are truly known and unknown, respectively. The prediction-selection disagreement is $\varepsilon_{\mathrm{con}} = \Pr_t(F(i) \neq F_\tau(i))$.*

Theorem 1 (Open-set risk bound with OOD refinement). *For the fixed representations and selection rule defined above, the model's open-set target risk satisfies*

$$\begin{aligned} R_t^{\mathrm{os}}(F) \leq (1 - \pi_{\mathrm{unk}}) \left[R_s(F^{\mathrm{kn}}) + \mathcal{D}_{\mathcal{H}} + \Lambda_{\mathrm{kn}} + \alpha_0 \right] \\ + \pi_{\mathrm{unk}} \beta_0 + m_{\mathrm{kn}}(\tau) - m_{\mathrm{unk}}(\tau) + \varepsilon_{\mathrm{con}}. \end{aligned} \tag{1}$$

The proof is provided in Appendix E. Theorem 1 highlights that OOD refinement contributes through the recovery of previously unidentified unknown nodes $m_{\mathrm{unk}}(\tau)$ at the cost of incorrectly rejecting known nodes $m_{\mathrm{kn}}(\tau)$. This motivates combining OOD evidence with prediction and clustering signals to guide selective alignment. The disagreement term $\varepsilon_{\mathrm{con}}$ further motivates incorporating the refined assignments into classifier training.

4 Methodology

Figure 2 illustrates the overall framework of OGSA. We first combine classifier predictions and prototype-initialized clustering to obtain initial target pseudo-labels (Section 4.1). Structure-aware OOD evidence then identifies additional unknown nodes among unresolved ones (Section 4.2). The refined assignments guide signed domain adversarial learning (Section 4.3) and target supervision (Section 4.4).

4.1 Prototype-Initialized Clustering

To obtain initial target assignments, we combine classifier predictions with clustering initialized from source class prototypes and candidate unknown target nodes. Let $\mathbf{p}_{r,i} = \mathrm{softmax}(\mathbf{z}_{r,i})$ denote the predicted class probabilities for node i in domain $r \in \{s, t\}$, with $p_{r,i,c}$ its c-th entry. We select candidate unknown nodes using two scores:

$$S_i^{(1)} = p_{t,i,K}, \quad S_i^{(2)} = 1 - \frac{\max_{0 \leq c < K} p_{t,i,c}}{1 - p_{t,i,K}}. \tag{2}$$

The first measures the predicted unknown class probability, while the second measures uncertainty among the known classes after renormalizing their probabilities. For a selection budget $1 \leq k \leq |V_t|$, let T_1 and T_2 contain the k highest-scoring target nodes under $S^{(1)}$ and $S^{(2)}$. Their intersection

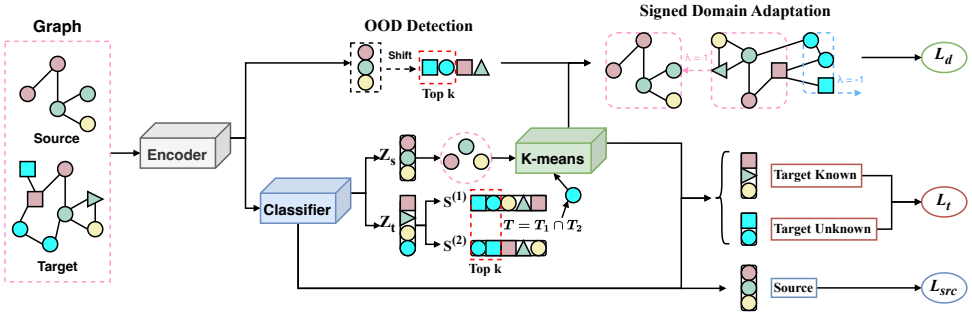


Figure 2: Overview of OGSA. The model first learns node representations via a graph encoder. Target nodes are then analyzed using uncertainty based selection and prototype-initialized clustering, followed by structure-aware OOD detection to identify unknown samples. Based on the inferred labels, a signed domain adversarial alignment is performed, where shared class samples are aligned with the source domain while detected unknown samples are repelled.

$T = T_1 \cap T_2$ provides the candidate unknown nodes. For each known class c, we initialize its prototype with the mean source representation $\boldsymbol{\mu}_c = |V_{s,c}|^{-1} \sum_{i \in V_{s,c}} \mathbf{h}_{s,i}$, where $V_{s,c} = \{i \in V_s : y_{s,i} = c\}$. The unknown prototype is initialized as $\boldsymbol{\mu}_K = |T|^{-1} \sum_{i \in T} \mathbf{h}_{t,i}$. We apply k-means clustering with $K+1$ clusters to the target representations, initialized with $\{\boldsymbol{\mu}_0, \ldots, \boldsymbol{\mu}_K\}$. Cluster indices follow the ordering of these initial prototypes. Let a_i denote the resulting cluster assignment for target node i. We assign $\hat{y}_{t,i}^{(0)} = a_i$ when $a_i = F(i)$ and set $\hat{y}_{t,i}^{(0)} = -1$ otherwise, marking unresolved nodes for OOD refinement.

4.2 OOD-Based Unknown Refinement

We then incorporate structural uncertainty using the GEBM framework (Fuchsgruber et al., 2024) to identify potential unknown nodes left unresolved by classifier-clustering agreement. It measures uncertainty by combining energy estimates from multiple structural levels. These complementary views capture both prediction confidence and structural consistency within the graph. Specifically, the overall structural uncertainty score for node i is defined as:

$$E_i = \log \left(\exp(\widehat{E}_I(i)) + \exp(\widehat{E}_L(i)) + \exp(\widehat{E}_G(i)) \right). \tag{3}$$

Detailed definitions of the energy terms are provided in Appendix B. We use E_i as the uncertainty score, where larger values indicate higher likelihood of being out-of-distribution. Among unresolved target nodes, we select the top-k nodes with the highest uncertainty scores to form a candidate set U and define the additional unknown set as $\Omega = T \cap U$. This restricts refinement to unresolved nodes supported by both prediction-based screening and OOD evidence. It instantiates the refinement in Theorem 1 with $\mathcal{C} = T \cap \{i : \hat{y}_{t,i}^{(0)} = -1\}$ and τ set to the k-th largest E_i among unresolved nodes, so that $\Omega = \Omega_\tau$. The refined pseudo-labels are $\hat{y}_{t,i} = K$ for $i \in \Omega$ and retain $\hat{y}_{t,i} = \hat{y}_{t,i}^{(0)}$ otherwise.

4.3 Signed Domain Adversarial Learning

We use the refined assignments to control the direction of domain adaptation through signed gradient reversal (Shen et al., 2025). Let $D_\phi(\mathbf{h})$ denote the probability that a representation belongs to the target domain. The discriminator minimizes the binary cross-entropy loss

$$\mathcal{L}_{\mathrm{d}} = -\frac{1}{|V_s| + |V_t|} \left[\sum_{i \in V_s} \log(1 - D_\phi(\mathbf{h}_{s,i})) + \sum_{i \in V_t} \log D_\phi(\mathbf{h}_{t,i}) \right]. \tag{4}$$

When updating the encoder, we reverse the domain loss gradients for source nodes and target nodes not assigned to the unknown class, so that the encoder maximizes $\mathcal{L}_{\mathrm{d}}$ on them and aligns them with the source domain. We preserve the gradient direction for target nodes with $\hat{y}_{t,i} = K$, so that the encoder minimizes $\mathcal{L}_{\mathrm{d}}$ on them and actively pushes their representations away from the source domain.

4.4 Training Procedure

Selective alignment requires identifying potential unknown nodes before adaptation begins, but source supervision alone provides no direct training signal for the unknown class. We therefore first pretrain the model with source supervision and auxiliary unknown class objectives to obtain initial predictions. During adaptation, these predictions are refined through clustering and OOD evidence to guide target supervision and signed domain adversarial learning.

Pretraining. We learn to distinguish known classes through the source cross-entropy loss $\mathcal{L}_{\text{src}} = -|V_s|^{-1} \sum_{i \in V_s} \log p_{s,i,(y_{s,i})}$. To introduce a training signal for the unknown output, we mask each source node's true class logit and encourage the unknown output over the remaining classes:

$$\mathcal{L}_{\text{unk}} = -\frac{1}{|V_s|} \sum_{i \in V_s} \log \frac{p_{s,i,K}}{1 - p_{s,i,(y_{s,i})}}. \tag{5}$$

For unlabeled target nodes, we use an adversarial binary cross-entropy objective (Saito et al., 2018; Shen et al., 2025):

$$\mathcal{L}_{\text{adv}} = -\frac{1}{2|V_t|} \sum_{i \in V_t} \left[\log p_{t,i,K} + \log(1 - p_{t,i,K}) \right]. \tag{6}$$

The classifier minimizes this loss toward an unknown probability of 0.5, while the encoder receives reversed gradients. The pretraining objective is $\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{src}} + \lambda_{\text{unk}} \mathcal{L}_{\text{unk}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}}$.

Adaptation. Let $\mathcal{P} = \{i \in V_t : \hat{y}_{t,i} \neq -1\}$ denote the target nodes with assigned pseudo-labels. Their classification loss is

$$\mathcal{L}_{\text{t}} = -\frac{1}{|\mathcal{P}|} \left[w_{\text{kn}} \sum_{\substack{i \in \mathcal{P} \\ \hat{y}_{t,i} < K}} \log p_{t,i,(\hat{y}_{t,i})} + \sum_{\substack{i \in \mathcal{P} \\ \hat{y}_{t,i} = K}} \log p_{t,i,K} \right], \tag{7}$$

where $w_{\text{kn}} > 0$ controls the relative weight of known class assignments. The overall adaptation objective is $\mathcal{L}_{\text{adapt}} = \mathcal{L}_{\text{src}} + \lambda_{\text{t}} \mathcal{L}_{\text{t}} + \lambda_{\text{d}} \mathcal{L}_{\text{d}}$, which we minimize using the signed gradient reversal rule in Section 4.3.

5 Experiments

5.1 Experimental Setup

Datasets. We evaluate OGSA on three benchmark groups with statistics summarized in Table 1. The citation benchmarks include five-class graphs DBLPv7 (D1), Citationv1 (C1), and ACMv9 (A1) (Liu et al., 2024b; Zhang & He, 2025), and nine-class graphs Citationv1 (C2), DBLPv4 (D2), and ACMv8 (A2) (Shen et al., 2025). Nodes represent papers and edges denote citations. For larger-scale temporal transfer, we use ogbn-arxiv (Hu et al., 2020), containing 40 paper categories and 128-dimensional features derived from titles and abstracts. Following Zhang & He (2025), we use three temporal domains: Arxiv I, II, and III.

Table 1: Dataset statistics.

Datasets	#Nodes	#Edges	#Feat	#Class
DBLPv7	5,484	8,117	6,775	5
Citationv1	8,935	15,098	6,775	5
ACMv9	9,360	15,556	6,775	5
Citationv1	9,737	14,054	7,786	9
DBLPv4	8,653	12,967	7,786	9
ACMv8	8,806	17,661	7,786	9
ogbn-arxiv	169,343	1,166,243	128	40

Baselines. We compare OGSA with 16 baselines from three complementary categories. (1) *Graph OOD Detection*: These methods evaluate unknown-node detection without explicit cross-domain alignment. We include GNNSafe (Wu et al., 2023b), which propagates energy scores over graph neighborhoods; GEBM (Fuchsgruber et al., 2024), which estimates uncertainty using multiscale graph energy; and DeGEM (Chen et al., 2025), GOLD (Wang et al., 2025), RSL (Yang et al., 2025), and EGonc (Zhang et al., 2024). (2) *Closed-set DA*: These methods evaluate cross-network adaptation without explicitly accounting for unknown target classes. We include UDAGCN (Wu et al., 2020a), A2GNN (Liu et al., 2024a), AdaGCN (Dai et al., 2023), SGDA (Qiao et al., 2023), and GRADE (Wu et al., 2023a). (3) *Open-set DA*: These methods are evaluated for known-class transfer and unknown-class recognition under domain shifts. We include DANCE (Saito et al., 2020),

Table 2: Average results (%) over the transfer tasks of each benchmark group. Bold and underline indicate the best and second-best results. Per-task results with standard deviations are reported in Appendix A.

Category	Method	Five-class citation			Nine-class citation			ogbn-arxiv		
		ACC	HS	AUC	ACC	HS	AUC	ACC	HS	AUC
Graph OOD Detection	GNNSafe	54.49	55.90	71.74	49.76	50.20	63.34	40.66	44.79	57.53
	DeGEM	53.11	34.42	69.64	43.20	36.78	57.42	34.25	32.88	51.26
	GOLD	40.04	36.12	66.44	52.84	46.60	66.71	38.36	41.51	56.01
	GEBM	58.56	61.14	75.46	55.43	<u>56.02</u>	67.35	40.35	43.94	58.24
	RSL	46.30	45.75	58.38	46.43	47.04	58.97	36.76	39.73	58.71
	EGonc	58.61	57.79	73.39	51.52	50.38	63.66	38.80	41.79	65.12
Closed-set DA	UDAGCN	41.81	38.51	55.99	50.68	50.74	60.78	46.21	50.89	64.45
	A2GNN	<u>63.19</u>	<u>65.00</u>	<u>76.62</u>	54.19	53.93	64.02	44.79	48.92	63.80
	AdaGCN	28.58	14.14	49.83	39.93	33.64	59.38	11.65	11.54	49.47
	SGDA	56.31	58.07	68.12	53.95	54.69	64.32	36.85	40.51	55.60
	GRADE	51.85	53.94	66.07	53.12	53.54	64.49	45.59	50.37	64.23
Open-set DA	DANCE	53.20	56.00	67.27	52.12	52.64	63.58	48.39	<u>52.70</u>	69.46
	PGL	30.37	16.70	59.76	40.15	16.23	56.87	21.89	10.55	58.32
	G2Pxy	59.74	61.82	75.84	49.61	45.09	64.23	36.29	37.91	61.39
	UAGA	59.80	58.98	71.55	<u>60.76</u>	54.19	<u>75.74</u>	48.06	43.04	65.27
	GraphRTA	59.74	61.73	76.36	52.71	51.55	68.77	<u>48.78</u>	48.17	<u>70.09</u>
OGSA		**74.37**	**75.78**	**93.16**	**65.47**	**65.91**	**79.17**	**53.56**	**55.20**	**77.19**

PGL (Luo et al., 2020), G2Pxy (Zhang et al., 2023), which generates proxy unknowns; UAGA (Shen et al., 2025), which excludes likely unknown nodes from domain alignment; and GraphRTA (Zhang & He, 2025), which jointly reprograms the target graph and source-trained model.

Implementation Details. Following prior studies on cross-network node classification (Zhang & He, 2025; Shen et al., 2025), we adopt the same open-set settings. For the five-class citation datasets, the first three classes are treated as known and the remaining two as unknown. For the nine-class citation datasets, the first five classes are known and the rest are unknown. For the ogbn-arxiv dataset, the first 30 classes are treated as known and the remaining classes as unknown. Our implementation is based on PyTorch and optimized using Adam. The learning rates of the pretraining and adaptation stages are searched within $[10^{-4}, 10^{-2}]$. The target supervision weight λ_{t} is searched within $[0.1, 2]$, the known class weight w_{kn} within $[0.01, 1]$, and λ_{adv} from $\{10^{-3}, 10^{-2}, 1\}$. We set $\lambda_{\mathrm{unk}} = 0.1$ and $\lambda_{\mathrm{d}} = 1$. The number of candidate nodes k is searched within $[1500, 4000]$ for citation networks and set to 10000 for ogbn-arxiv. The pretraining stage runs for 30 epochs on citation networks and 40 epochs on ogbn-arxiv, followed by 150 epochs of domain adaptation. Following Shen et al. (2025), the GAT encoder uses 8 attention heads with 32 dimensions each, while the node classifier uses 2 heads. The batch size is 2048 for all datasets.

Evaluation Metrics. We evaluate performance using overall accuracy (ACC), H-score (HS), and the area under the receiver operating characteristic curve (AUC). ACC measures classification accuracy over all target nodes, with all unknown classes merged into label K. HS is the harmonic mean: $\mathrm{HS} = 2\,\mathrm{Acc}_{\mathrm{kn}}\mathrm{Acc}_{\mathrm{unk}}/(\mathrm{Acc}_{\mathrm{kn}} + \mathrm{Acc}_{\mathrm{unk}})$, where $\mathrm{Acc}_{\mathrm{kn}}$ is the accuracy on known-class target nodes and $\mathrm{Acc}_{\mathrm{unk}}$ is the proportion of unknown target nodes correctly identified as unknown. AUC measures how well the unknownness scores distinguish unknown from known target nodes, treating unknown nodes as the positive class. We run each method with five random seeds and report the mean and standard deviation of each metric.

5.2 Main Results

Table 2 reports the average results over each benchmark group, and the per-task results are provided in Appendix A. OGSA achieves the best ACC and HS on all 15 transfer tasks and the best AUC on 14 of them. Averaged over tasks, OGSA achieves absolute ACC and HS improvements over the strongest baseline of each task of 9.83% and 10.11% on the five-class citation benchmarks, 4.52% and 7.64% on the nine-class citation benchmarks, and 3.85% and 2.50% on ogbn-arxiv.

Graph OOD detectors identify unknown nodes without cross-domain alignment. OGSA consistently outperforms them, indicating that OOD evidence alone is insufficient since distributional deviation also arises from domain shifts within shared classes. Closed-set DA methods align target nodes without accounting for unknown classes. A2GNN is often the second-best method on the citation benchmarks, yet its much lower AUC than OGSA reflects the risk of aligning unknown nodes with source classes. Open-set DA methods address both aspects but rely primarily on prediction-related signals to identify unknown nodes. OGSA clearly outperforms UAGA despite sharing its signed alignment mechanism, highlighting the value of OOD evidence for unknown identification.

The relative strength of the baseline categories varies across benchmarks. On the citation benchmarks, the strongest baseline in HS is a closed-set DA method or a graph OOD detector on 10 of the 12 tasks, whereas open-set DA methods lead on only two. This suggests that prediction-related signals alone provide unreliable unknown estimates under these shifts. On ogbn-arxiv, which involves 40 classes and temporal shifts, open-set DA methods are the strongest baselines on all three tasks. Graph OOD detectors perform the worst, indicating that alignment becomes critical when the shift is large. In contrast to these baselines whose effectiveness depends on the benchmark, OGSA performs best across all benchmark groups since it combines OOD evidence for unknown identification with selective alignment for shared classes.

5.3 Ablation Studies

Component Analysis. Table 3 evaluates the contribution of each component of OGSA on the five-class citation benchmarks. (1) Removing pretraining reduces the average ACC and HS by 5.32% and 4.98%. Without pretraining, the model has not learned to predict the unknown class when adaptation begins. The nodes it initially flags as likely unknown are therefore less reliable. Since these nodes initialize the unknown cluster and constrain OOD refinement, the resulting pseudo-labels become less accurate. (2) Removing OOD refinement causes the largest degradation, reducing the average ACC and HS by 6.29% and 6.50%. Relying on OOD evidence alone also reduces them by 5.45% and 4.21%. Without OOD evidence, unknown nodes that prediction and clustering fail to resolve are never recovered and are aligned with the source domain. With OOD evidence alone, known nodes that deviate from the source distribution due to domain shift are mistaken for unknown nodes. Combining both sources recovers missed unknown nodes while avoiding the false rejection of known nodes. (3) Aligning all target nodes with the source domain reduces the average ACC and HS by 3.15% and 3.25%. Once unknown nodes are aligned, their representations move toward known class regions and become harder to reject. Pushing them away from the source domain prevents this negative transfer.

Robustness to Pseudo-Label Quality. The component analysis shows that OGSA benefits from accurate pseudo-labels. This raises the question of whether OGSA depends heavily on the quality of early pseudo-labels, since errors in them could be reinforced when the same pseudo-labels guide supervision and alignment. To examine this issue, we track the quality of the pseudo-labels used in each epoch in Appendix D.2. As adaptation proceeds, OGSA steadily improves the precision of both known and unknown pseudo-labels. The fraction of unknown nodes mislabeled as known drops substantially while the fraction of known nodes mislabeled as unknown remains low. More importantly, OGSA recovers the pseudo-label quality in the early stage of adaptation even without pretraining when the initial pseudo-labels are highly inaccurate. This demonstrates that OGSA corrects pseudo-labeling errors rather than amplifying them. Note that the variant without pretraining shows lower error rates than OGSA on some tasks. This is because fewer nodes receive pseudo-labels in this variant rather than because they are identified more accurately, as reflected by its lower precision.

Choice of OOD Scores. The component analysis shows that OOD refinement is the most critical component of OGSA. A natural question is whether this benefit depends on the specific OOD score. We therefore replace the GEBM score used to construct U with its partial variants and with the Mahalanobis distance in Appendix D.1. All variants achieve an average HS between 75.44% and 75.78%, which is about 6% higher than the variant without OOD refinement. This indicates that the improvement of OGSA stems from incorporating distributional evidence into unknown identification rather than from a particular choice of OOD score.

Table 3: Ablation study on five-class citation datasets. Results are mean (standard deviation) (%). Bold indicates the best means.

Model	C1→A1		D1→A1		A1→C1		C1→D1		D1→C1		A1→D1	
	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS
OGSA	**71.98 (0.87)**	**75.37 (0.77)**	**68.76 (1.14)**	**70.53 (1.19)**	**81.43 (0.17)**	**80.53 (0.33)**	**74.54 (0.48)**	**76.87 (0.77)**	**75.94 (1.34)**	**77.04 (1.25)**	**73.54 (0.45)**	**74.36 (0.77)**
w/o pretraining	62.20 (7.18)	64.98 (8.42)	62.25 (1.98)	65.68 (2.14)	73.33 (1.23)	76.50 (1.07)	68.15 (3.80)	68.51 (4.46)	74.97 (2.53)	75.34 (3.32)	73.37 (1.79)	73.83 (4.70)
w/o OOD refinement	70.59 (1.23)	74.09 (1.22)	62.81 (1.70)	66.25 (1.87)	69.51 (5.60)	69.11 (4.69)	63.37 (1.10)	66.75 (2.54)	68.78 (5.70)	66.78 (6.03)	73.38 (2.17)	72.75 (1.27)
OOD refinement only	64.26 (3.83)	67.59 (4.28)	59.99 (3.39)	60.35 (4.44)	75.78 (2.50)	78.74 (1.91)	69.28 (1.59)	73.66 (1.12)	72.41 (1.26)	75.54 (0.93)	71.74 (1.20)	73.58 (1.05)
w/o signed alignment	63.77 (1.51)	65.37 (2.22)	66.58 (2.65)	67.93 (2.01)	81.38 (0.21)	79.86 (0.29)	73.10 (0.51)	76.39 (0.64)	72.10 (1.07)	74.22 (1.02)	70.39 (3.17)	71.42 (1.23)



(a) GEBM (b) A2GNN (c) UAGA (d) GraphRTA (e) OGSA

Figure 3: t-SNE visualization of node embeddings on the Citation dataset (C1→D1), where red points denote unknown classes.

Representation Analysis. We further examine whether OGSA preserves the separation between known and unknown nodes in the learned representations. Appendix D.3 repeats the distance analysis in Figure 1 with OGSA. After adaptation, unknown nodes remain farther from the source known class distributions than known nodes in the lower confidence groups, where OOD refinement mainly operates. Figure 3 further visualizes the target representations on C1→D1. GEBM identifies unknown nodes without cross-domain alignment and therefore leaves known classes poorly separated. A2GNN aligns all target nodes with the source domain so that unknown nodes are scattered among known class clusters. UAGA and GraphRTA partially separate unknown nodes but still show noticeable overlap with known classes. In contrast, OGSA forms a compact unknown cluster that is clearly separated from the known classes. Both analyses indicate that OGSA aligns shared classes while keeping unknown nodes distinguishable from them.

Parameter Analysis. Appendix D.4 shows that removing $\mathcal{L}_{\text{unk}}$ or $\mathcal{L}_{\text{adv}}$ alone reduces the average HS by 1.14% and 1.25%, while removing both reduces it by 3.36%, indicating that the two objectives complement each other in initializing unknown prediction. Removing $\mathcal{L}_{\text{d}}$ also consistently lowers performance. Appendix D.5 varies the number of unknown classes. With a single unknown class, unknown nodes provides limited evidence for learning unknownness and makes the unknown accuracy in HS more sensitive to errors. With three unknown classes, unknown nodes span several semantically different classes and overlap more with known regions, making them harder to separate as a whole. Appendix D.6 further shows that OGSA remains stable over wide ranges of most hyperparameters.

6 Conclusion

We propose OGSA, an OOD-guided selective alignment framework for open-set cross-network node classification. Existing open-set methods rely primarily on prediction-related signals to identify unknown nodes, whereas OGSA further exploits distributional deviation from the source domain as complementary evidence. The refined unknown assignments then guide target supervision and signed domain adversarial learning to align shared classes and push unknown nodes away. Experiments show that OGSA achieves the best ACC and HS on all 15 transfer tasks. Ablation studies further confirm that both sources of evidence are necessary and that iterative refinement corrects pseudo-labeling errors rather than amplifying them. These findings suggest that distributional evidence deserves explicit consideration in open-set graph domain adaptation.

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

We provide the implementation as supplementary material and will make the code publicly available upon acceptance. Section 4 describes the model and training procedure, and Algorithm 1 summarizes the full pipeline. Dataset statistics, open-set settings, hyperparameter search ranges, and evaluation protocols are provided in Table 1 and Section 5.1, and the compute resources are described in Appendix C. The complete proof of Theorem 1 is included in Appendix E.

References

Jushuo Chen, Feifei Dai, Xiaoyan Gu, Jiang Zhou, Bo Li, and Weipinng Wang. Universal domain adaptive network embedding for node classification. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 4022–4030, 2023.

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In *International conference on machine learning*, pp. 1725–1735. PMLR, 2020.

Yuhan Chen, Yihong Luo, Yifan Song, Pengwen Dai, Jing Tang, and Xiaochun Cao. Decoupled graph energy-based model for node out-of-distribution detection on heterophilic graphs. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 78270–78292, 2025.

Quanyu Dai, Xiao-Ming Wu, Jiaren Xiao, Xiao Shen, and Dan Wang. Graph transfer learning via adversarial domain adaptation with graph convolution. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4908–4922, 2023.

Dominik Fuchsgruber, Tom Wollschläger, and Stephan Günnemann. Energy-based epistemic uncertainty for graph neural networks. *Advances in Neural Information Processing Systems*, 37: 34378–34428, 2024.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

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(1):723–773, 2012.

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

Russell Hart, Linlin Yu, Yifei Lou, and Feng Chen. Improvements on uncertainty quantification for node classification via distance based regularization. *Advances in Neural Information Processing Systems*, 36:55454–55478, 2023.

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.

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

Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31, 2018.

Guohao Li, Matthias Muller, Ali Thabet, and Bernard Ghanem. Deepgcns: Can gcns go as deep as cnns? In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9267–9276, 2019.

Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017.

Meihan Liu, Zeyu Fang, Zhen Zhang, Ming Gu, Sheng Zhou, Xin Wang, and Jiajun Bu. Rethinking propagation for unsupervised graph domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 13963–13971, 2024a.

Meihan Liu, Zhen Zhang, Jiachen Tang, Jiajun Bu, Bingsheng He, and Sheng Zhou. Revisiting, benchmarking and understanding unsupervised graph domain adaptation. *Advances in neural information processing systems*, 37:89408–89436, 2024b.

Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in neural information processing systems*, 33:21464–21475, 2020.

Yadan Luo, Zijian Wang, Zi Huang, and Mahsa Baktashmotlagh. Progressive graph learning for open-set domain adaptation. In *International Conference on Machine Learning*, pp. 6468–6478. PMLR, 2020.

Longfei Ma, Yiyou Sun, Kaize Ding, Zemin Liu, and Fei Wu. Revisiting score propagation in graph out-of-distribution detection. *Advances in Neural Information Processing Systems*, 37: 4341–4373, 2024.

Ziyue Qiao, Xiao Luo, Meng Xiao, Hao Dong, Yuanchun Zhou, and Hui Xiong. Semi-supervised domain adaptation in graph transfer learning. *arXiv preprint arXiv:2309.10773*, 2023.

Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 153–168, 2018.

Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. *Advances in neural information processing systems*, 33:16282–16292, 2020.

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, 2020a.

Xiao Shen, Quanyu Dai, Sitong Mao, Fu-lai Chung, and Kup-Sze Choi. Network together: Node classification via cross-network deep network embedding. *IEEE Transactions on Neural Networks and Learning Systems*, 32(5):1935–1948, 2020b.

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.

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

Danny Wang, Ruihong Qiu, Guangdong Bai, and Zi Huang. Gold: Graph out-of-distribution detection via implicit adversarial latent generation. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 45297–45326, 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.

Zirui Wang, Zihang Dai, Barnabás Póczos, and Jaime Carbonell. Characterizing and avoiding negative transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11293–11302, 2019.

Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *International conference on machine learning*, pp. 6861–6871. Pmlr, 2019.

Jun Wu, Jingrui He, and Elizabeth Ainsworth. Non-iid transfer learning on graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 10342–10350, 2023a.

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, 2020a.

Man Wu, Shirui Pan, and Xingquan Zhu. Openwgl: Open-world graph learning. In *2020 IEEE international conference on data mining (icdm)*, pp. 681–690. IEEE, 2020b.

Man Wu, Shirui Pan, and Xingquan Zhu. Attraction and repulsion: Unsupervised domain adaptive graph contrastive learning network. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(5):1079–1091, 2022.

Qitian Wu, Yiting Chen, Chenxiao Yang, and Junchi Yan. Energy-based out-of-distribution detection for graph neural networks. *arXiv preprint arXiv:2302.02914*, 2023b.

Shenzhi Yang, Junbo Zhao, Sharon Li, Shouqing Yang, Dingyu Yang, Xiaofang Zhang, and Haobo Wang. Harnessing feature resonance under arbitrary target alignment for out-of-distribution node detection. 38, Main Conference:162201–162238, 2025. doi: 10.52202/085713-5417.

Shuiqiao Yang, Borui Cai, Taotao Cai, Xiangyu Song, Jiaojiao Jiang, Bing Li, and Jianxin Li. Robust cross-network node classification via constrained graph mutual information. *Knowledge-Based Systems*, 257:109852, 2022.

Qin Zhang, Zelin Shi, Xiaolin Zhang, Xiaojun Chen, Philippe Fournier-Viger, and Shirui Pan. $\mathcal{G}^2$Pxy: Generative open-set node classification on graphs with proxy unknowns. *arXiv preprint arXiv:2308.05463*, 2023.

Qin Zhang, Zelin Shi, Shirui Pan, Junyang Chen, Huisi Wu, and Xiaojun Chen. Egonc: Energy-based open-set node classification with substitute unknowns. *Advances in Neural Information Processing Systems*, 37:66147–66177, 2024.

Zhen Zhang and Bingsheng He. Towards unsupervised open-set graph domain adaptation via dual reprogramming. *arXiv preprint arXiv:2510.18363*, 2025.

A FULL RESULTS

Tables 4 and 5 report the per-task results on the citation and ogbn-arxiv benchmarks.

Table 4: Open-set cross-network node classification on citation networks. Results are mean (standard deviation) (%). Bold and underline indicate the best and second-best means.

Task	Metric	Graph OOD Detection						Closed-set DA					Open-set DA					OGSA
		GNNSafe	DeGEM	GOLD	GEBM	RSL	EGonc	UDAGCN	A2GNN	AdaGCN	SGDA	GRADE	DANCE	PGL	G2Pxy	UAGA	GraphRTA	
C1→D1	ACC	57.08 (2.05)	61.37 (2.32)	29.84 (1.98)	62.43 (1.89)	49.34 (0.89)	61.70 (0.42)	54.84 (2.64)	61.27 (0.24)	26.95 (4.16)	59.42 (1.70)	55.80 (0.31)	54.57 (2.91)	28.59 (0.22)	62.76 (0.45)	52.13 (2.85)	62.92 (0.78)	**74.54 (0.48)**
	HS	59.44 (3.45)	45.46 (22.57)	21.72 (4.35)	66.00 (1.31)	52.76 (0.85)	58.12 (1.52)	54.30 (3.28)	64.02 (0.27)	15.76 (12.64)	61.76 (2.65)	58.99 (0.43)	58.34 (5.91)	18.33 (0.33)	63.94 (0.79)	63.72 (2.62)	63.92 (1.01)	**76.87 (0.77)**
	AUC	72.59 (3.79)	79.31 (4.43)	73.57 (2.63)	80.51 (0.23)	63.77 (1.35)	75.05 (0.33)	64.16 (2.74)	74.78 (0.18)	52.19 (5.61)	72.99 (3.13)	70.29 (0.34)	68.94 (8.62)	62.20 (1.06)	77.67 (0.55)	77.90 (2.35)	78.74 (0.54)	**92.55 (0.33)**
C1→A1	ACC	56.87 (0.50)	50.01 (5.08)	32.37 (4.65)	59.59 (0.56)	45.69 (1.44)	57.40 (0.37)	44.34 (5.37)	59.72 (0.51)	30.18 (7.54)	52.49 (1.39)	49.90 (0.34)	53.22 (0.88)	32.70 (0.58)	58.92 (0.42)	60.91 (0.96)	58.75 (1.17)	**71.98 (0.87)**
	HS	56.96 (1.58)	41.80 (15.67)	15.82 (14.57)	62.46 (0.36)	45.77 (3.45)	56.77 (0.40)	45.08 (7.75)	60.75 (1.07)	11.93 (16.31)	51.85 (1.77)	52.58 (0.37)	55.32 (1.56)	17.20 (0.86)	60.45 (0.50)	58.33 (1.05)	60.32 (1.97)	**75.37 (0.77)**
	AUC	73.92 (0.77)	74.08 (6.98)	56.74 (11.11)	78.15 (0.14)	58.45 (0.50)	72.76 (0.87)	62.20 (4.56)	73.72 (0.94)	47.31 (12.43)	63.74 (1.10)	65.40 (0.72)	66.29 (1.02)	60.29 (4.24)	75.75 (0.99)	68.98 (2.42)	76.93 (2.07)	**94.01 (0.57)**
D1→C1	ACC	52.12 (1.31)	56.78 (3.03)	42.36 (4.69)	58.27 (0.13)	41.84 (1.27)	59.34 (0.82)	31.15 (4.70)	69.27 (0.16)	25.74 (0.39)	56.66 (1.27)	47.99 (0.90)	53.56 (1.66)	30.65 (0.85)	59.48 (0.18)	60.41 (0.37)	58.31 (1.08)	**75.94 (1.34)**
	HS	55.21 (1.59)	40.51 (22.56)	41.79 (8.55)	61.04 (0.34)	43.59 (2.04)	60.95 (0.86)	27.88 (13.75)	69.74 (0.15)	17.11 (11.22)	59.17 (1.21)	50.10 (1.07)	56.12 (1.83)	15.93 (1.11)	62.48 (0.15)	52.26 (0.55)	59.98 (1.47)	**77.04 (1.25)**
	AUC	70.05 (1.43)	78.18 (4.73)	65.07 (2.78)	73.69 (0.78)	51.85 (3.89)	73.27 (0.99)	41.08 (5.74)	77.97 (0.24)	45.12 (4.49)	66.71 (1.40)	61.42 (0.56)	66.49 (2.14)	59.80 (3.15)	74.38 (0.52)	70.48 (0.75)	71.30 (1.91)	**93.64 (0.59)**
D1→A1	ACC	52.12 (0.50)	47.72 (2.61)	43.63 (4.28)	56.04 (0.39)	43.45 (3.66)	55.21 (1.10)	35.33 (4.16)	65.94 (0.08)	31.50 (2.85)	52.35 (1.28)	50.25 (0.60)	53.26 (0.58)	32.21 (1.07)	56.94 (0.68)	60.86 (0.52)	57.11 (1.59)	**68.76 (1.14)**
	HS	52.81 (1.24)	32.44 (17.95)	42.51 (8.12)	58.61 (0.31)	35.69 (12.92)	57.79 (1.09)	27.22 (14.22)	66.84 (0.15)	15.52 (17.25)	53.98 (2.94)	52.38 (0.75)	56.12 (0.60)	16.20 (1.54)	59.97 (0.72)	62.13 (0.82)	59.61 (1.37)	**70.53 (1.19)**
	AUC	71.53 (0.44)	69.87 (3.31)	69.89 (1.87)	74.98 (0.28)	54.24 (6.15)	74.89 (0.98)	55.33 (3.22)	82.45 (0.09)	50.06 (9.08)	65.87 (3.22)	68.56 (0.60)	70.60 (3.12)	56.51 (4.44)	77.85 (0.39)	71.30 (1.43)	77.71 (1.03)	**91.16 (0.92)**
A1→C1	ACC	51.34 (1.48)	50.99 (4.91)	39.94 (4.33)	59.94 (0.48)	47.59 (2.89)	61.20 (0.58)	36.34 (8.01)	65.95 (0.23)	26.28 (0.94)	58.09 (2.55)	52.22 (1.11)	58.24 (1.05)	30.85 (1.16)	61.72 (0.48)	62.23 (0.34)	62.25 (1.53)	**81.43 (0.17)**
	HS	54.50 (1.59)	30.08 (23.78)	40.17 (6.46)	61.74 (0.80)	45.65 (4.40)	58.80 (0.50)	29.75 (16.15)	67.06 (0.17)	15.65 (12.82)	61.08 (2.85)	52.25 (1.52)	59.39 (1.73)	16.16 (1.45)	63.23 (0.64)	59.14 (0.33)	64.63 (1.56)	**80.53 (0.33)**
	AUC	69.25 (0.87)	62.15 (12.71)	61.26 (4.64)	73.32 (0.84)	58.00 (5.35)	72.05 (0.55)	53.77 (3.44)	75.54 (0.19)	52.03 (5.24)	69.55 (4.98)	60.93 (1.32)	66.59 (2.16)	60.46 (4.68)	73.57 (0.99)	70.77 (0.88)	75.43 (2.58)	**95.89 (0.04)**
A1→D1	ACC	57.43 (0.61)	51.78 (1.73)	52.12 (3.87)	55.11 (0.46)	49.91 (1.06)	56.79 (0.75)	48.85 (8.45)	57.00 (0.07)	30.86 (3.28)	58.88 (1.10)	54.92 (0.65)	46.36 (2.24)	27.20 (1.02)	58.63 (0.37)	62.23 (1.19)	59.08 (1.83)	**73.54 (0.45)**
	HS	56.47 (1.69)	16.22 (21.68)	54.71 (2.81)	56.97 (3.77)	51.05 (1.28)	54.31 (0.88)	46.85 (9.71)	61.56 (0.09)	8.88 (13.02)	60.57 (0.92)	57.34 (1.04)	50.73 (2.39)	16.36 (1.28)	60.84 (0.57)	58.29 (1.57)	61.93 (0.91)	**74.36 (0.77)**
	AUC	73.11 (1.21)	54.25 (19.67)	72.11 (1.70)	72.12 (4.67)	63.94 (1.20)	72.30 (0.59)	59.40 (9.74)	75.24 (0.19)	52.25 (3.26)	69.87 (0.86)	69.80 (0.99)	64.70 (2.04)	59.30 (5.81)	75.80 (0.49)	69.89 (1.92)	78.04 (1.16)	**91.72 (0.19)**
C2→D2	ACC	51.20 (0.89)	50.31 (6.55)	48.89 (0.67)	57.62 (0.29)	47.81 (0.19)	48.79 (1.67)	45.66 (9.74)	53.95 (0.44)	35.68 (10.86)	56.85 (1.02)	51.44 (1.48)	52.69 (1.18)	44.10 (0.67)	50.51 (0.47)	62.36 (0.28)	49.46 (4.13)	**65.33 (1.57)**
	HS	51.32 (0.88)	39.50 (18.12)	43.50 (1.26)	57.94 (0.23)	47.83 (0.36)	42.86 (3.01)	42.91 (10.08)	52.83 (0.37)	23.72 (20.24)	57.01 (0.94)	51.62 (1.62)	52.37 (1.69)	17.68 (0.97)	50.79 (0.45)	56.21 (0.63)	47.71 (5.91)	**65.23 (1.64)**
	AUC	59.51 (1.07)	65.65 (9.23)	57.57 (1.36)	66.05 (0.55)	58.56 (0.16)	56.81 (2.11)	56.16 (9.01)	56.15 (0.68)	54.09 (3.74)	66.32 (0.73)	57.38 (1.89)	61.66 (1.63)	57.42 (1.97)	57.38 (0.76)	73.32 (0.99)	65.23 (2.27)	**75.46 (2.70)**
C2→A2	ACC	48.22 (1.09)	42.18 (4.71)	51.47 (0.88)	52.96 (0.80)	42.16 (1.26)	47.95 (0.56)	49.82 (1.67)	42.53 (1.19)	39.63 (7.19)	51.84 (1.29)	44.83 (3.36)	50.09 (2.32)	37.34 (0.49)	46.52 (0.35)	64.87 (0.17)	50.69 (3.03)	**65.86 (0.85)**
	HS	49.32 (0.95)	40.54 (8.17)	43.86 (5.36)	54.18 (0.93)	43.22 (1.31)	46.52 (1.60)	50.51 (1.99)	43.40 (1.54)	30.91 (13.17)	52.76 (1.82)	45.18 (3.77)	50.54 (2.91)	15.40 (0.82)	46.72 (0.32)	61.28 (0.43)	49.49 (5.11)	**67.27 (0.91)**
	AUC	63.72 (1.17)	58.86 (6.76)	64.09 (1.27)	67.33 (1.00)	56.70 (1.86)	58.94 (0.73)	63.13 (1.90)	54.03 (1.42)	58.94 (2.86)	64.83 (1.23)	56.49 (4.04)	61.10 (3.88)	54.86 (7.05)	58.64 (0.38)	**83.12 (0.23)**	60.61 (1.46)	82.04 (1.15)
D2→C2	ACC	49.69 (0.87)	43.34 (3.95)	55.85 (1.06)	55.33 (0.39)	47.94 (0.13)	54.81 (0.79)	51.46 (0.71)	56.61 (0.47)	40.76 (1.39)	55.82 (0.58)	57.36 (0.27)	52.64 (1.02)	40.82 (0.85)	54.78 (0.33)	59.95 (0.56)	51.19 (2.39)	**63.98 (1.83)**
	HS	50.46 (1.02)	35.99 (10.59)	54.11 (3.40)	56.02 (0.38)	48.75 (0.12)	55.29 (1.11)	52.26 (0.85)	57.52 (0.48)	28.62 (6.40)	56.74 (0.67)	57.62 (0.33)	53.43 (1.09)	18.62 (1.26)	51.29 (0.65)	52.52 (0.99)	52.82 (3.07)	**63.99 (1.40)**
	AUC	63.91 (1.41)	57.60 (5.21)	67.39 (0.86)	65.37 (0.59)	61.62 (0.15)	67.58 (0.62)	60.81 (1.12)	69.21 (0.60)	63.55 (2.31)	65.15 (0.42)	69.53 (0.27)	63.36 (1.62)	60.33 (1.77)	69.85 (0.35)	72.38 (0.95)	71.39 (0.56)	**76.12 (0.66)**
D2→A2	ACC	51.14 (0.36)	41.49 (3.94)	51.29 (5.01)	53.60 (0.41)	43.99 (0.29)	52.16 (0.33)	49.79 (1.57)	51.41 (0.29)	42.02 (6.46)	51.88 (2.67)	52.51 (0.32)	50.37 (0.72)	37.68 (0.81)	47.70 (0.32)	56.85 (0.65)	52.94 (1.17)	**62.03 (0.48)**
	HS	52.38 (0.42)	35.70 (9.53)	37.49 (23.18)	54.59 (0.48)	44.90 (0.20)	53.28 (0.43)	50.50 (1.86)	51.63 (0.20)	39.73 (7.33)	53.12 (2.76)	53.77 (0.33)	51.52 (0.76)	16.41 (1.06)	45.55 (0.53)	50.29 (1.40)	49.80 (2.05)	**63.43 (0.49)**
	AUC	69.80 (0.69)	56.06 (6.78)	74.15 (1.16)	68.41 (0.47)	57.93 (0.30)	69.07 (0.85)	61.31 (1.88)	66.00 (0.22)	58.64 (2.95)	65.12 (1.41)	69.57 (0.42)	65.23 (0.78)	59.29 (0.72)	67.47 (0.65)	76.81 (1.51)	68.91 (0.58)	**78.85 (0.54)**
A2→C2	ACC	48.97 (0.78)	42.01 (5.34)	55.94 (1.13)	55.82 (0.65)	48.74 (0.13)	52.97 (0.74)	51.97 (0.24)	61.22 (0.61)	32.69 (11.46)	54.01 (1.06)	56.10 (0.88)	53.12 (1.19)	39.10 (1.17)	46.66 (0.66)	62.26 (1.30)	52.82 (1.86)	**67.02 (0.26)**
	HS	49.02 (0.80)	37.18 (11.35)	54.87 (2.56)	56.02 (1.04)	49.56 (0.12)	53.17 (0.59)	52.88 (0.24)	59.31 (1.18)	31.20 (12.90)	54.86 (1.05)	56.77 (1.10)	53.90 (1.30)	15.81 (1.82)	33.78 (1.69)	55.13 (0.80)	54.53 (2.28)	**67.40 (0.24)**
	AUC	62.86 (1.60)	54.68 (9.15)	69.57 (1.16)	68.13 (1.23)	60.66 (0.15)	65.42 (1.18)	60.99 (0.62)	71.08 (0.93)	59.93 (2.51)	62.70 (1.42)	67.16 (0.77)	64.27 (1.45)	55.54 (4.15)	66.40 (0.82)	77.16 (0.18)	70.33 (0.40)	**79.56 (0.25)**
A2→D2	ACC	49.33 (1.55)	39.84 (8.83)	53.60 (4.34)	57.26 (1.08)	47.95 (0.28)	52.43 (1.90)	55.40 (1.12)	59.39 (1.13)	48.77 (5.04)	53.32 (0.22)	56.48 (1.34)	53.81 (1.43)	41.87 (1.16)	51.50 (0.53)	58.26 (1.07)	59.13 (0.68)	**68.58 (1.10)**
	HS	48.71 (1.89)	31.76 (16.98)	45.76 (15.44)	57.36 (0.95)	47.95 (0.39)	51.17 (3.54)	55.41 (0.91)	58.86 (1.24)	47.67 (5.06)	53.65 (0.20)	56.27 (1.17)	54.10 (1.34)	13.47 (1.64)	42.42 (1.29)	49.73 (0.78)	54.93 (2.36)	**68.13 (1.45)**
	AUC	60.23 (1.86)	51.69 (14.72)	67.48 (1.30)	68.80 (1.50)	58.35 (0.54)	64.12 (2.04)	62.30 (1.96)	67.62 (1.35)	61.15 (5.11)	61.81 (0.71)	66.79 (1.72)	65.88 (1.21)	53.75 (5.55)	65.67 (0.59)	71.68 (1.65)	76.12 (0.65)	**83.00 (1.64)**

Table 5: Open-set cross-network node classification on ogbn-arxiv. Results are mean (standard deviation) (%). Bold and underline indicate the best and second-best means.

Task	Metric	Graph OOD Detection						Closed-set DA					Open-set DA					OGSA
		GNNSafe	DeGEM	GOLD	GEBM	RSL	EGonc	UDAGCN	A2GNN	AdaGCN	SGDA	GRADE	DANCE	PGL	G2Pxy	UAGA	GraphRTA	
ArxivI→ArxivII	ACC	41.54 (0.30)	35.14 (5.90)	39.86 (0.96)	41.62 (0.28)	36.86 (2.51)	40.71 (2.12)	47.30 (0.34)	46.19 (0.19)	11.07 (1.34)	36.56 (2.48)	47.22 (0.16)	47.75 (0.78)	22.76 (0.17)	37.91 (1.16)	49.93 (2.83)	48.73 (1.43)	**55.10 (0.82)**
	HS	45.92 (0.44)	33.91 (12.30)	41.41 (1.86)	45.53 (0.51)	40.00 (3.68)	43.79 (3.20)	52.06 (0.40)	49.41 (0.24)	9.65 (5.01)	39.88 (2.53)	51.61 (0.27)	52.50 (0.97)	10.70 (0.39)	39.46 (2.27)	45.80 (6.54)	50.89 (4.20)	**56.80 (1.31)**
	AUC	58.49 (1.15)	55.21 (13.85)	57.04 (2.25)	60.14 (0.35)	59.42 (4.41)	67.07 (1.14)	65.18 (0.74)	63.78 (0.22)	49.87 (0.33)	53.13 (2.14)	64.96 (0.44)	68.25 (1.54)	61.35 (0.94)	63.57 (1.26)	66.43 (6.53)	70.70 (4.31)	**79.35 (2.50)**
ArxivI→ArxivIII	ACC	40.10 (0.42)	35.96 (6.93)	35.50 (0.97)	38.11 (0.28)	34.36 (2.93)	34.76 (1.64)	43.34 (0.51)	42.14 (0.16)	4.35 (3.42)	32.68 (1.12)	42.72 (0.18)	46.74 (0.59)	21.50 (0.25)	33.51 (1.40)	44.23 (2.80)	45.16 (0.72)	**50.36 (0.22)**
	HS	43.85 (0.55)	31.37 (16.46)	38.93 (1.52)	42.70 (0.46)	37.63 (4.68)	35.98 (2.63)	47.70 (0.63)	46.78 (0.23)	2.95 (1.94)	36.44 (1.02)	47.61 (0.23)	50.93 (1.23)	10.13 (0.58)	33.69 (2.85)	35.68 (7.20)	44.63 (2.88)	**52.11 (0.28)**
	AUC	56.73 (1.21)	51.62 (14.46)	54.55 (3.08)	57.63 (0.52)	59.30 (5.87)	62.91 (2.49)	61.68 (0.57)	62.16 (0.37)	50.07 (0.41)	53.34 (0.38)	62.25 (0.27)	68.83 (2.39)	57.74 (2.06)	61.33 (1.23)	58.55 (6.35)	63.78 (2.40)	**73.21 (0.68)**
ArxivII→ArxivIII	ACC	40.33 (0.97)	31.64 (6.99)	39.71 (0.60)	41.32 (0.27)	39.06 (0.38)	40.94 (0.33)	47.99 (0.63)	46.04 (0.23)	19.53 (0.14)	41.31 (0.98)	46.83 (0.08)	50.68 (0.52)	21.42 (0.34)	37.45 (0.63)	50.01 (0.92)	52.46 (1.80)	**55.21 (0.24)**
	HS	44.59 (0.88)	33.36 (9.50)	44.18 (0.78)	43.58 (0.54)	41.57 (0.46)	45.61 (0.47)	52.92 (0.73)	50.56 (0.27)	22.02 (0.17)	45.21 (1.36)	51.88 (0.10)	54.67 (0.98)	10.83 (0.50)	40.58 (1.19)	47.64 (2.53)	49.00 (4.80)	**56.70 (0.39)**
	AUC	57.37 (1.18)	46.96 (10.84)	56.44 (1.51)	56.95 (0.69)	57.40 (1.16)	65.37 (0.51)	66.48 (1.23)	65.46 (0.36)	48.46 (1.95)	60.32 (2.14)	65.49 (0.39)	71.31 (1.34)	55.86 (1.56)	59.26 (1.08)	70.82 (2.38)	75.80 (4.43)	**79.02 (0.94)**

B DETAILS OF GEBM-BASED OOD SCORES

We provide the detailed formulation of the structure-aware OOD score based on GEBM (Fuchsgruber et al., 2024). Given the logits $\mathbf{z}_{t,i}$ of target node i, the class-conditional energy is $\mathcal{E}(i,c) = -z_{t,i,c}$. To mitigate overconfidence, a Gaussian regularization term is added:

$$\widehat{\mathcal{E}}(i,c) = \mathcal{E}(i,c) - \gamma \log \mathcal{N}(\mathbf{z}_{t,i} \mid \boldsymbol{m}_c, \mathbf{S}_c), \tag{8}$$

where $\boldsymbol{m}_c$ and $\mathbf{S}_c$ are the mean and covariance of source logits of class c, and $\gamma > 0$ controls the regularization strength. Let $\mathbf{P} = \left(\eta\mathbf{I} + (1-\eta)\mathbf{A}_t\mathbf{D}_t^{-1}\right)^M$ denote the graph propagation operator, where $\mathbf{D}_t$ is the degree matrix of $\mathbf{A}_t$, $\eta \in [0,1]$ is the teleport probability, and M is the number of propagation steps. The three energy terms in Eq. 3 are defined as follows.

Independent energy:

$$\widehat{E}_I(i) = -\log \sum_{c=0}^{K} \exp\big(-\widehat{\mathcal{E}}(i,c)\big). \tag{9}$$

Local energy:

$$\widehat{E}_L(i) = -\log \sum_{c=0}^{K} \exp\Big[\mathbf{P}\big(-\widehat{\mathcal{E}}(\cdot, c)\big)\Big]_i. \tag{10}$$

Group energy:

$$\widehat{E}_G(i) = -\Big[\mathbf{P}\Big(\log \sum_{c=0}^{K} \exp\big(-\widehat{\mathcal{E}}(\cdot, c)\big)\Big)\Big]_i. \tag{11}$$

These terms capture uncertainty from individual predictions, local structural consistency, and neighborhood-level aggregation of uncertainty, respectively.

C Compute Resources

All experiments are conducted on a high-performance computing cluster using a single NVIDIA H100 GPU with 80GB HBM3 memory. The system is based on an x86_64 architecture with up to 64 CPU cores. We use PyTorch 2.3.1 with CUDA 12.1 for implementation.

D Additional Experiments

D.1 OOD Scores

Table 6 replaces the GEBM score used to construct U with its partial variants and the Mahalanobis distance. All distribution-based scores achieve comparable performance, indicating that OGSA is robust to the choice of OOD score and that its gains stem from incorporating distributional evidence itself.

Table 6: Ablation study of OOD scores on five-class citation datasets. Results are mean (standard deviation) (%). Bold indicates the best means.

Model	C1 → A1		D1 → A1		A1 → C1		C1 → D1		D1 → C1		A1 → D1	
	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS
OGSA	**71.98 (0.87)**	**75.37 (0.77)**	68.76 (1.14)	70.53 (1.19)	**81.43 (0.17)**	**80.53 (0.33)**	74.54 (0.48)	76.87 (0.77)	75.94 (1.34)	77.04 (1.25)	73.54 (0.45)	74.36 (0.77)
w/o $\widehat{E}_G$	70.65 (1.41)	74.17 (1.36)	68.63 (0.73)	70.51 (0.78)	81.27 (0.24)	80.31 (0.27)	74.29 (0.67)	76.47 (0.85)	76.03 (1.81)	77.31 (1.72)	73.48 (0.53)	**74.71 (0.34)**
$\widehat{E}_I$ only	70.45 (1.43)	74.02 (1.40)	67.73 (1.48)	69.83 (1.10)	81.07 (0.21)	80.09 (0.41)	74.39 (0.43)	76.64 (0.61)	75.99 (1.61)	77.42 (1.57)	**73.60 (0.40)**	74.66 (0.45)
Mahalanobis	69.98 (1.74)	73.02 (2.62)	**69.11 (0.73)**	**70.72 (1.15)**	81.13 (0.35)	80.02 (0.50)	**74.84 (0.51)**	**77.22 (0.70)**	**77.09 (1.73)**	**78.57 (1.31)**	73.60 (0.45)	74.62 (0.81)

D.2 Robustness to Pseudo-Label Quality

Figure 4 tracks the quality of the target pseudo-labels used for target supervision and signed alignment in each epoch. For OGSA, the fraction of known nodes mislabeled as unknown remains low and stable. The fraction of unknown nodes mislabeled as known decreases on most tasks. Iterative refinement thus corrects pseudo-labeling errors rather than amplifying them. Without pretraining, pseudo-labels are much less accurate at the beginning of adaptation but recover quickly as training proceeds. This shows that OGSA does not collapse under poor initial pseudo-labels. Nevertheless, a gap to OGSA remains on most tasks at the end of training, indicating that pretraining improves not only the starting point but also the final pseudo-label quality. On some tasks, the variant without pretraining mislabels fewer unknown nodes as known. This reflects lower coverage rather than better identification, since fewer nodes receive pseudo-labels and their precision remains lower than that of OGSA.

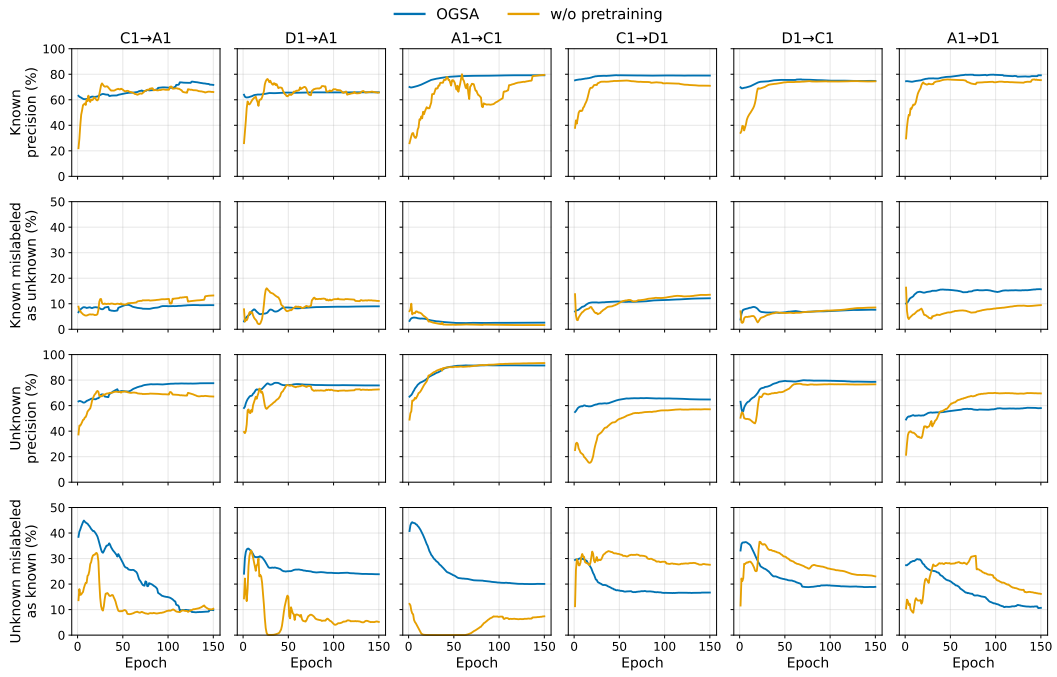
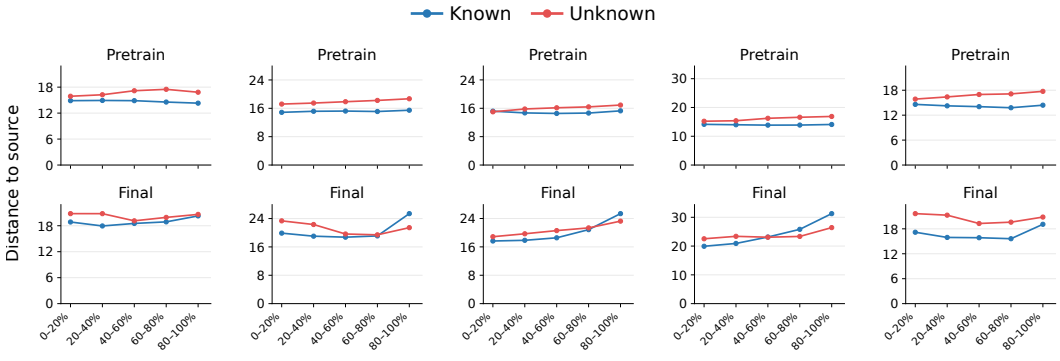


Figure 4: Pseudo-label quality during adaptation on five-class citation datasets, averaged over five seeds. Known and unknown precision are computed over nodes assigned known and unknown pseudo-labels. Known mislabeled as unknown and unknown mislabeled as known are computed over all truly known and truly unknown target nodes.

D.3 Distributional Separation after Adaptation

Figure 5 repeats the analysis in Figure 1 with representations learned by OGSA. After adaptation, unknown nodes remain farther from the source known class distributions than known nodes in the lower confidence groups across all tasks. These low-confidence nodes are the ones most likely to be left unresolved by prediction and clustering and are therefore the main targets of OOD refinement. OGSA thus preserves distributional evidence where it is most needed for unknown identification.



(a) C1→D1 (b) D1→C1 (c) C1→A1 (d) A1→D1 (e) A1→C1

Figure 5: Mean Mahalanobis distance of known and unknown target nodes to the source known class distributions across confidence quantile groups for OGSA, after pretraining (top) and after adaptation (bottom).

D.4 Auxiliary Objectives

Table 7 evaluates the auxiliary pretraining objectives and the domain adversarial loss. Removing either $\mathcal{L}_{\text{unk}}$ or $\mathcal{L}_{\text{adv}}$ reduces performance on most tasks, and removing both leads to the largest drop. Removing $\mathcal{L}_{\text{d}}$ also lowers ACC and HS on all six tasks.

Table 7: Ablation study of auxiliary objectives on five-class citation datasets. Results are mean (standard deviation) (%). Bold indicates the best means.

Model	C1→A1		D1→A1		A1→C1		C1→D1		D1→C1		A1→D1	
	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS
OGSA	**71.98 (0.87)**	**75.37 (0.77)**	**68.76 (1.14)**	**70.53 (1.19)**	**81.43 (0.17)**	**80.53 (0.33)**	**74.54 (0.48)**	**76.87 (0.77)**	**75.94 (1.34)**	**77.04 (1.25)**	**73.54 (0.45)**	**74.36 (0.77)**
w/o $\mathcal{L}_{\text{unk}}$	71.48 (0.74)	74.69 (0.61)	66.40 (1.73)	68.56 (1.59)	81.20 (0.23)	80.30 (0.34)	74.27 (0.25)	76.68 (0.40)	74.04 (1.05)	75.51 (1.05)	73.37 (0.24)	72.11 (0.62)
w/o $\mathcal{L}_{\text{adv}}$	68.84 (2.11)	71.43 (2.78)	67.03 (1.63)	68.54 (1.54)	81.39 (0.14)	80.41 (0.19)	74.35 (0.45)	76.73 (0.66)	74.47 (2.63)	75.88 (2.49)	73.32 (0.67)	74.21 (0.84)
w/o $\mathcal{L}_{\text{unk}}$, $\mathcal{L}_{\text{adv}}$	66.49 (6.59)	68.54 (8.34)	61.01 (2.78)	62.50 (3.03)	81.24 (0.17)	80.36 (0.27)	74.06 (0.42)	76.51 (0.62)	71.85 (1.65)	73.48 (1.33)	72.71 (0.76)	73.14 (0.48)
w/o $\mathcal{L}_{\text{d}}$	66.54 (1.38)	70.16 (1.32)	64.34 (0.55)	67.88 (0.55)	80.56 (0.21)	79.59 (0.35)	73.82 (0.40)	76.59 (0.50)	71.95 (1.66)	73.78 (1.48)	70.34 (0.35)	74.00 (0.38)

D.5 Number of Unknown Classes

Table 8 varies the number of unknown classes C_{unk} on the five-class citation datasets, where the remaining $5 - C_{\text{unk}}$ classes are treated as known.

Table 8: Effect of the number of unknown classes C_{unk} on five-class citation datasets. Results are mean (standard deviation) (%). $C_{\text{unk}} = 2$ is the default setting.

C_{unk}	C1→A1		D1→A1		A1→C1		C1→D1		D1→C1		A1→D1	
	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS	ACC	HS
1	63.70 (2.39)	67.04 (5.11)	62.30 (0.51)	67.81 (0.71)	74.77 (1.29)	77.95 (1.02)	68.49 (1.39)	74.84 (1.13)	72.16 (0.58)	76.32 (1.41)	71.71 (0.55)	76.33 (0.27)
2	71.98 (0.87)	75.37 (0.77)	68.76 (1.14)	70.53 (1.19)	81.43 (0.17)	80.53 (0.33)	74.54 (0.48)	76.87 (0.77)	75.94 (1.34)	77.04 (1.25)	73.54 (0.45)	74.36 (0.77)
3	64.03 (7.34)	63.81 (7.47)	67.33 (1.26)	63.43 (5.76)	73.05 (2.02)	71.61 (3.03)	65.66 (2.03)	63.75 (3.37)	71.44 (0.61)	70.66 (1.19)	63.15 (0.37)	60.90 (0.40)

D.6 Hyperparameter Sensitivity

Figure 6 reports the sensitivity of OGSA to its hyperparameters. Performance remains stable over wide ranges of λ_{unk}, λ_{t}, w_{kn}, λ_{d}, and k, while large values of λ_{adv} degrade performance.

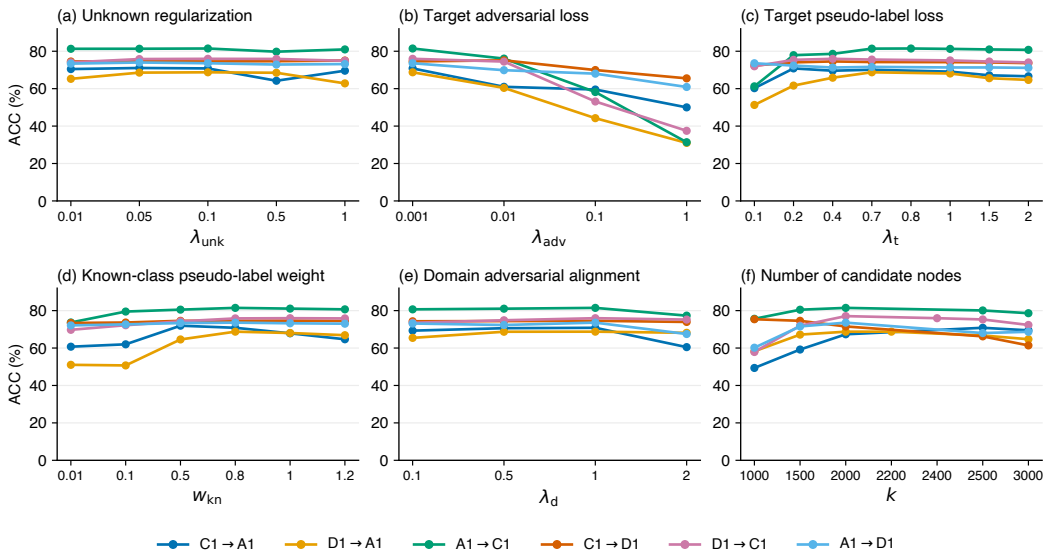


Figure 6: Sensitivity of OGSA to hyperparameters on five-class citation datasets.

E Proof of Theorem 1

Probabilities and expectations indexed by s and t are taken over a node drawn uniformly from the finite node set of the corresponding graph throughout the proof. The assumption $\pi_{\mathrm{unk}} \in (0,1)$ ensures that conditioning on $\widetilde{Y} < K$ and on $\widetilde{Y} = K$ is well defined.

Lemma 1 (Triangle inequality)**.** *For any labels a, b, and c, $\ell(a,c) \le \ell(a,b) + \ell(b,c)$.*

Proof. If $a \neq c$, then $a \neq b$ or $b \neq c$, so the right-hand side is at least one. Otherwise the left-hand side is zero.

Lemma 2 (Error on unknown nodes)**.** *The auxiliary predictor satisfies*

$$\Pr_t\big(F_\tau(i) \neq \widetilde{Y},\, \widetilde{Y} = K\big) = \pi_{\mathrm{unk}}\beta_0 - m_{\mathrm{unk}}(\tau). \tag{12}$$

Proof. When $b_i^\tau = 0$, we have $F_\tau(i) = F^{\mathrm{kn}}(i) < K$. When $b_i^\tau = 1$, we have $F_\tau(i) = K$. Hence for a node with $\widetilde{Y} = K$, the event $F_\tau(i) \neq \widetilde{Y}$ coincides with $b_i^\tau = 0$. By definition, $b_i^\tau = 0$ holds if and only if $b_i^0 = 0$ and $i \notin \Omega_\tau$. Since $\Omega_\tau \subseteq \mathcal{C} \subseteq \{i : b_i^0 = 0\}$, the event $\{b_i^0 = 0\}$ is the disjoint union of $\{b_i^0 = 0,\, i \notin \Omega_\tau\}$ and $\{i \in \Omega_\tau\}$. Therefore,

$$\begin{aligned}
\Pr_t\big(F_\tau(i) \neq \widetilde{Y},\, \widetilde{Y} = K\big) &= \Pr_t\big(b_i^0 = 0,\, \widetilde{Y} = K\big) - \Pr_t\big(i \in \Omega_\tau,\, \widetilde{Y} = K\big) \\
&= \Pr_t(\widetilde{Y} = K)\, \Pr_t\big(b_i^0 = 0 \mid \widetilde{Y} = K\big) - m_{\mathrm{unk}}(\tau) \\
&= \pi_{\mathrm{unk}}\beta_0 - m_{\mathrm{unk}}(\tau).
\end{aligned}$$

Lemma 3 (Error on known nodes)**.** *The auxiliary predictor satisfies*

$$\Pr_t\big(F_\tau(i) \neq \widetilde{Y},\, \widetilde{Y} < K\big) \le (1 - \pi_{\mathrm{unk}})\big[\alpha_0 + R_{t,\mathrm{kn}}(F^{\mathrm{kn}})\big] + m_{\mathrm{kn}}(\tau). \tag{13}$$

Proof. For a node with $\widetilde{Y} < K$, we have $\widetilde{Y} = Y$. If $b_i^\tau = 1$, then $F_\tau(i) = K \neq Y$ and the prediction is wrong. If $b_i^\tau = 0$, then $F_\tau(i) = F^{\mathrm{kn}}(i)$ and the prediction is wrong exactly when $F^{\mathrm{kn}}(i) \neq Y$. Hence

$$\begin{aligned}
\Pr_t\big(F_\tau(i) \neq \widetilde{Y},\, \widetilde{Y} < K\big) &= \Pr_t\big(b_i^\tau = 1,\, \widetilde{Y} < K\big) + \Pr_t\big(b_i^\tau = 0,\, F^{\mathrm{kn}}(i) \neq Y,\, \widetilde{Y} < K\big) \\
&\le \Pr_t\big(b_i^\tau = 1,\, \widetilde{Y} < K\big) + \Pr_t\big(F^{\mathrm{kn}}(i) \neq Y,\, \widetilde{Y} < K\big).
\end{aligned}$$

For the first term, $b_i^\tau = 1$ holds if and only if $b_i^0 = 1$ or $i \in \Omega_\tau$, and these two events are disjoint because $\Omega_\tau \subseteq \{i : b_i^0 = 0\}$. Thus

$$\Pr_t\big(b_i^\tau = 1,\, \widetilde{Y} < K\big) = \Pr_t\big(b_i^0 = 1,\, \widetilde{Y} < K\big) + \Pr_t\big(i \in \Omega_\tau,\, \widetilde{Y} < K\big) = (1 - \pi_{\mathrm{unk}})\alpha_0 + m_{\mathrm{kn}}(\tau).$$

For the second term, by the definition of the conditional risk,

$$\Pr_t\big(F^{\mathrm{kn}}(i) \neq Y,\, \widetilde{Y} < K\big) = \Pr_t(\widetilde{Y} < K)\, \mathbb{E}_t\big[\ell(F^{\mathrm{kn}}(\mathbf{h}), Y) \mid \widetilde{Y} < K\big] = (1 - \pi_{\mathrm{unk}})R_{t,\mathrm{kn}}(F^{\mathrm{kn}}).$$

Combining the two terms yields Eq. 13.

Lemma 4 (Known class risk under domain shift)**.** *The known class predictor satisfies*

$$R_{t,\mathrm{kn}}(F^{\mathrm{kn}}) \le R_s(F^{\mathrm{kn}}) + \mathcal{D}_{\mathcal{H}} + \Lambda_{\mathrm{kn}}. \tag{14}$$

Proof. Fix any $q \in \mathcal{H}$. Applying Lemma 1 with $a = F^{\mathrm{kn}}(\mathbf{h})$, $b = q(\mathbf{h})$ and $c = Y$, then taking the expectation over target known class nodes gives

$$R_{t,\mathrm{kn}}(F^{\mathrm{kn}}) \le \mathbb{E}_{t,\mathrm{kn}}\, \ell\big(F^{\mathrm{kn}}(\mathbf{h}), q(\mathbf{h})\big) + R_{t,\mathrm{kn}}(q).$$

Since both F^{kn} and q belong to $\mathcal{H}$, the definition of $\mathcal{D}_{\mathcal{H}}$ implies

$$\mathbb{E}_{t,\mathrm{kn}}\,\ell\big(F^{\mathrm{kn}}(\mathbf{h}), q(\mathbf{h})\big) \le \mathbb{E}_{s}\,\ell\big(F^{\mathrm{kn}}(\mathbf{h}), q(\mathbf{h})\big) + \mathcal{D}_{\mathcal{H}}.$$

Applying Lemma 1 again on the source graph where all labels lie in $\mathcal{Y}_s$, gives

$$\mathbb{E}_{s}\,\ell\big(F^{\mathrm{kn}}(\mathbf{h}), q(\mathbf{h})\big) \le R_s(F^{\mathrm{kn}}) + R_s(q).$$

Combining the three inequalities yields

$$R_{t,\mathrm{kn}}(F^{\mathrm{kn}}) \le R_s(F^{\mathrm{kn}}) + \mathcal{D}_{\mathcal{H}} + R_s(q) + R_{t,\mathrm{kn}}(q).$$

Because this holds for every $q \in \mathcal{H}$, taking the infimum over q gives Eq. 14.

Theorem 1 For the fixed representations and selection rule defined above, the model's open-set target risk satisfies

$$\begin{aligned} R_t^{\mathrm{os}}(F) \le (1-\pi_{\mathrm{unk}})\left[R_s(F^{\mathrm{kn}}) + \mathcal{D}_{\mathcal{H}} + \Lambda_{\mathrm{kn}} + \alpha_0\right] \\ + \pi_{\mathrm{unk}}\beta_0 + m_{\mathrm{kn}}(\tau) - m_{\mathrm{unk}}(\tau) + \varepsilon_{\mathrm{con}}. \end{aligned} \tag{15}$$

Proof. Applying Lemma 1 with $a = F(i)$, $b = F_\tau(i)$, and $c = \widetilde{Y}$ and taking the expectation over target nodes gives

$$R_t^{\mathrm{os}}(F) \le \Pr_t\big(F(i) \neq F_\tau(i)\big) + \Pr_t\big(F_\tau(i) \neq \widetilde{Y}\big) = \varepsilon_{\mathrm{con}} + \Pr_t\big(F_\tau(i) \neq \widetilde{Y}\big). \tag{16}$$

We split the second term according to whether the node belongs to a known or an unknown class:

$$\Pr_t\big(F_\tau(i) \neq \widetilde{Y}\big) = \Pr_t\big(F_\tau(i) \neq \widetilde{Y},\, \widetilde{Y} < K\big) + \Pr_t\big(F_\tau(i) \neq \widetilde{Y},\, \widetilde{Y} = K\big).$$

Applying Lemmas 2, 3, and 4 to the two terms, we obtain

$$\begin{aligned} \Pr_t\big(F_\tau(i) \neq \widetilde{Y}\big) &\le (1-\pi_{\mathrm{unk}})\big[\alpha_0 + R_{t,\mathrm{kn}}(F^{\mathrm{kn}})\big] + m_{\mathrm{kn}}(\tau) + \pi_{\mathrm{unk}}\beta_0 - m_{\mathrm{unk}}(\tau) \\ &\le (1-\pi_{\mathrm{unk}})\big[R_s(F^{\mathrm{kn}}) + \mathcal{D}_{\mathcal{H}} + \Lambda_{\mathrm{kn}} + \alpha_0\big] + \pi_{\mathrm{unk}}\beta_0 + m_{\mathrm{kn}}(\tau) - m_{\mathrm{unk}}(\tau). \end{aligned}$$

Substituting this into Eq. 16 yields Eq. 15.

F ALGORITHM

Algorithm 1 OGSA

Input: Source graph $\mathcal{G}_s$ with labels $\mathbf{Y}_s$, target graph $\mathcal{G}_t$;
epochs N_{pre}, N_{ada}; budget k; GEBM refit interval Δ;
weights $\lambda_{\text{unk}}, \lambda_{\text{adv}}, \lambda_{\text{t}}, \lambda_{\text{d}}, w_{\text{kn}}$.
Output: Target predictions $F(i)$ for $i \in V_t$.

Initialize g_θ, f_ψ, and D_ϕ.
for $e = 1, \ldots, N_{\text{pre}}$ **do** ▷ Pretraining
  Compute $\mathcal{L}_{\text{src}}$, $\mathcal{L}_{\text{unk}}$ (Eq. 5), and $\mathcal{L}_{\text{adv}}$ (Eq. 6).
  Update θ, ψ by minimizing $\mathcal{L}_{\text{pre}}$ with $\nabla\mathcal{L}_{\text{adv}}$ reversed for g_θ.
end for
for $e = 1, \ldots, N_{\text{ada}}$ **do** ▷ Adaptation
  Compute $\mathbf{H}_s$, $\mathbf{H}_t$, $\mathbf{p}_{t,i}$, and $F(i)$.
  Select $T = T_1 \cap T_2$ by $S_i^{(1)}$ and $S_i^{(2)}$ (Eq. 2).
  Run k-means from $\{\boldsymbol{\mu}_0, \ldots, \boldsymbol{\mu}_K\}$ to obtain a_i. ▷ Clustering
  Set $\hat{y}_{t,i}^{(0)} = a_i$ if $a_i = F(i)$, and $\hat{y}_{t,i}^{(0)} = -1$ otherwise.
  if $e = 1$ **or** $e \bmod \Delta = 0$ **then**
    Refit GEBM (Appendix B).
  end if
  Compute E_i (Eq. 3).
  Let U be the top-k unresolved nodes by E_i and $\Omega = T \cap U$. ▷ OOD refinement
  Set $\hat{y}_{t,i} = K$ for $i \in \Omega$, and $\hat{y}_{t,i} = \hat{y}_{t,i}^{(0)}$ otherwise.
  for each mini-batch **do** ▷ Signed alignment
    Compute $\mathcal{L}_{\text{src}}$, $\mathcal{L}_{\text{t}}$ (Eq. 7), and $\mathcal{L}_{\text{d}}$ (Eq. 4).
    Update ϕ by minimizing $\mathcal{L}_{\text{d}}$.
    Update θ, ψ by minimizing $\mathcal{L}_{\text{adapt}}$ with signed gradient reversal.
  end for
end for
return $F(i)$ for all $i \in V_t$.