Prioritizing Potential Wetland Areas via Region-to-Region Knowledge Transfer and Adaptive Propagation

Yoonhyuk Choi
Arizona State University
Computer Engineering
Tempe, USA
ychoi139@asu.edu

Reepal Shah
Tulane University
ByWater Institute
New Orleans, USA
rshah3@tulane.edu

John Sabo
Tulane University
ByWater Institute & River-Coastal Science and Engineering
New Orleans, USA
jsabo1@tulane.edu

Huan Liu
Arizona State University
Computer Engineering
Tempe, USA
huanliu@asu.edu

K. Selçuk Candan
Arizona State University
Computer Engineering
Tempe, USA
candan@asu.edu

Abstract—Wetlands are important to communities, offering benefits ranging from water purification, and flood protection to recreation and tourism. Therefore, identifying and prioritizing potential wetland areas is a critical decision problem. While data-driven solutions are feasible, this is complicated by significant data sparsity due to the low proportion of wetlands (3-6%) in many areas of interest in the southwestern US. This makes it hard to develop data-driven models that can help guide the identification of additional wetland areas. To solve this limitation, we propose two strategies: (1) knowledge transfer from regions with rich wetlands (such as the Eastern US) to regions with sparser wetlands (such as the Southwestern area). , and (2) spatial data enrichment strategy that relies on an adaptive propagation mechanism. This mechanism differentiates between node pairs that have positive and negative impacts on each other for Graph Neural Networks (GNNs). We conduct rigorous experiments to substantiate our proposed method's effectiveness, robustness, and scalability compared to state-of-the-art baselines. Additionally, an ablation study demonstrates that each module is essential in prioritizing potential wetlands.

Index Terms—Wetland identification, Transfer learning, Domain disentanglement, Adaptive propagation, Data sparsity

I. Introduction

Wetlands, characterized by water at or near the soil surface for most of the year, play several crucial roles within an ecosystem. Specifically, they provide habitat for various plant and animal species, improving water quality by filtering pollutants and storing carbon that mitigates climate change [2]. Furthermore, wetlands are also important to communities for improving water quality by filtering pollutants and flood protection [3]. Despite their value to the ecosystem, wetlands face challenges from competition with agriculture, urban development, and climate change. These highlight the need for both the conservation of existing wetlands and the identification of potential new wetland areas.
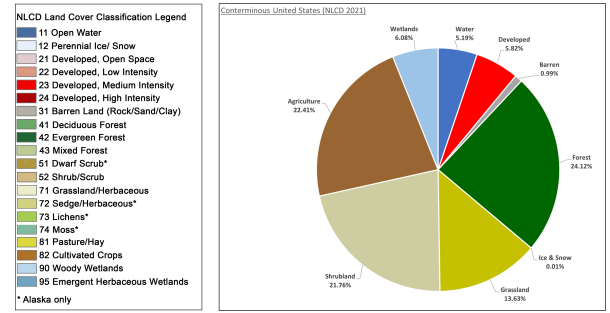

Fig. 1: Natural Land Cover Dataset (NLCD [1] provides 20 categories of land use, including wetlands (types 90 and 95). The data reveals that only 6% of the land in the US is wetland

A. Challenges and Solutions

A major challenge in developing new wetlands, particularly in areas where they are essential for water conservation, is the arid climate and local soil features that make water retention difficult. As shown in Figure 1, wetlands only account for 6% of land cover in the US and the share of wetlands gets even lower in dryer regions. The sparsity of wetlands makes it hard to develop data-driven models that can help guide the identification of new wetland areas. In particular, given the large imbalance between the wetland and non-wetland regions, naively developed models can suffer from overfitting and bias issues. This issue is further complicated considering that regions with few wetlands are not necessarily suitable for wetland development, making it particularly challenging to develop data-driven models with sparse positive samples.

In this paper, we argue that this *data sparsity* challenge can be addressed by relying on two complementary techniques:

- *Region-to-region knowledge transfer:* We can transfer knowledge from regions with rich wetlands (such as the Eastern US) to regions where wetlands are sparse (such as the Southwestern US with few wetlands). However, this solution suffers from **global context incompatibility** *challenge* – since the source and target regions are likely to differ significantly in climate, soil characteristics, population distribution, and land use. Moreover, the expected outcomes like water storage and flood prevention from the wetlands make it difficult to transfer knowledge from the source context to the target.
- *Adaptive knowledge transfer within local regions:* We complement this with a spatial data enrichment strategy that transfers useful information among the spatial cells, with different characteristics, in the same vicinity. This solution also encounters a form of **local context incompatibility,** where information from nearby cells should be considered differently: while some cell pairs *positively* affect each other, others have a *negative* impact. Hence, it's crucial to identify and consider this aspect when implementing knowledge transfer between local regions.

In summary, we propose a novel method of identifying ***Pot**ential wetlands via **T**ransfer learning and **A**daptive propagation (PoTA)* to tackle the data sparsity challenge in wetland prioritization problem, which is equipped with two complementary knowledge transfer techniques:

- We suggest a long-range, region-to-region knowledge transfer from locations rich in wetlands to sparser ones using domain disentanglement
- Coupled with adaptive feature propagation within individual regions, we employ an enhanced message-passing scheme between heterophilic neighbors.
- We provide theoretical analyses of the effectiveness of the proposed components in solving data sparsity issues.

II. RELATED WORK

A. Wetland Identification

Wetlands are known to occur due to reasons such as permanent or seasonal inundation or soil saturation [4]. Due to the value of their resources, prioritizing potential wetlands has become an important topic [2]. Generally, the formation of wetlands is known to be closely related to soil characteristics [5]. Therefore, various methods have been proposed for wetland identification, such as collecting geographic data based on sensing [6] or finding indicators [7] necessary for wetland formation. Based on these associations, the most recent approach presented is solving the problem using deep learning techniques on the physically-informed data [8].

B. Knowledge Transfer

Knowledge transfer (a.k.a transfer learning) refers to using additional data (from the source domain) to solve data sparsity when training data is lacking in the target domain. Generally, this method utilizes the common contexts that exist in both domains, which is widely used in text classification [9], digit recognition [10], and recommender systems [11]. The

common context means some information like images, digits, ratings, and so on. In our case, the natural features like soil and drainage features from different regions could be used for transfer learning. Here, we focus on the cross-domain recommendation schemes [12], [13], which suggest capturing common knowledge from both domains. The foundational concept called domain adaptation [14], [15] has been proposed, which captures domain-shareable features (common knowledge) through adversarial training. However, several recent studies have revealed that the common features without specific guidance may not always be helpful to the target domain [16], [17]. This becomes more severe as the discrepancy between domains (e.g., different categories) increases, leading to the proposal of domain disentanglement techniques [18], [19] to address this issue.

C. Adaptive Propagation

Information propagation [20] is a widely used technique in graph theory. In detail, the mechanism of Graph Neural Networks (GNNs) is node embedding and message-passing (propagation) [21], which has the advantage of using the adjacent nodes for the prediction. However, message-passing algorithms may fail to perform well under heterophilic graphs [22], where most edges connect two nodes with different labels. To conquer this problem, several studies suggested finding these connections [23] by measuring the difference of nodes (e.g., attention) or by utilizing remote nodes with high similarity (non-local aggregation). In addition, [24] proposes ego-neighbor separation for message-passing, [25] generates a compatibility matrix, [26] further utilizes neighbors (non-local propagation), and [27] figures out the path-level pattern. As another branch, the mechanism of adaptive propagation [28], and choosing appropriate architectures [29] have been recently proposed. The core concept of adaptive propagation is that they determine the sign of edges [30], [31] before applying a message-passing scheme, which can be either positive or negative.

III. PRELIMINARIES

In this section, we present the preliminaries, including useful notations that will be used in this paper. We first introduce the natural features available for wetland prioritization. Then, we formulate the wetland prioritization problem.

A. Land Cover and Surface Features

In this paper, we primarily rely on the Natural Land Cover Dataset (**NLCD 2021**[1]) visualized in Figure 1. The data set includes the land cover type (20 categories, at $30m \times 30m$ resolution) from 2001 to 2021. This data set is complemented with soil type information, from **SSURGO**[2] data set, and drainage information from *Height Above the Nearest Drainage* (**HAND**) data set[3]. As described in Table I, the SSURGO

---

[1]https://www.indianamap.org/maps/INMap::nlcd-land-cover-2021/about
[2]https://www.nrcs.usda.gov/resources/data-and-reports/gridded-soil-survey-geographic-gssurgo-database
[3]https://registry.opendata.aws/glo-30-hand/

TABLE I: Surface features collected from the SSURGO and HAND data sets

Data	Description	Notation
SSURGO	Soil characteristic	s_1
	Slope gradient (min, max)	s_2, s_3
	Water table depth	s_4
	Flood frequency (min, max)	s_5, s_6
	Available water storage	$s_7 - s_{10}$
	Drainage class	s_{11}
	Hydrologic group	$s_{12} - s_{15}$
	Others (position, SOC, and etc)	$s_{16} - s_{34}$
HAND	Height above the nearest drainage	s_{35}
NLCD	Wetland label	s_{36}

feature consists of soil characteristics, slope gradient, water table depth, available water storage, and so on. Note that these two data sets have different resolutions from the NLCD data set – for instance, the SSURGO data set has a finer $10m \times 10m$ resolution. Therefore, we integrated the three data sets by matching the nearest coordinates. The final aligned data sets have $30m \times 30m$ (9 grids) resolution.

B. Problem Formulation

The wetland prioritization task can be formulated as follows. As mentioned above, for each target region (R_τ), we have land cover (NLCD), soil (SSURGO), and drainage (HAND) data. Specifically, for each cell c in the range, we have a wetland label $w_c \in \{0, 1\}$ (wetland if $w_c = 1$), multi-dimensional soil feature vector s_c, and drainage flag value $h_c \in (0, 1)$. Coupled with these, we further assume an adjacency matrix $\mathcal{G}_\tau$, which describes the neighborhood relationship among the cells in the region. Given these, we define the *wetland classification* problem as follows:

(Wetland Classification) Given the set of tuples $D_\tau = \{(c, \mathcal{G}_c, s_c, h_c) \mid c \in R_\tau\}$, the *wetland classification* problem aims to find a mapping function $f_\tau(c, \mathcal{G}_c, s_c, h_c) \to w_c$.

The careful reader would note that the wetland classification problem, in and of itself, will not enable wetland identification and prioritization tasks as the only thing a perfect model with 100% accuracy would provide is an explanation of the current wetlands in the given region, rather than recommending new wetlands. Therefore, we seek models with *high recalls* (i.e., accurately explaining existing wetlands in the region), while not necessarily having perfect Accuracy. For instance, a cell c in the target region R_τ where the model predicts a wetland, despite the absence of one currently, may serve as a candidate for new wetland development.

As discussed before, our goal is to improve the wetland identification and prioritization task for target regions with sparse wetlands by relying on information obtained from source regions with denser wetlands. Therefore, we present a slightly modified version of the problem as follows:

(Wetland Classification with Knowledge Transfer) Given a source region, R_s and a target region R_τ and the associated sets of tuples D_s and D_τ, the *wetland classification with*

knowledge transfer problem aims to find a mapping function $f_{s,\tau}(c, \mathcal{G}_c, s_c, h_c) \to w_c$, where $c \in R_\tau$.

As before we seek models with *high recalls* for the target region. High Accuracy is preferred, with the understanding that a cell $c \in R_\tau$ for which $w_c = 0$ and $f_{s,\tau}(c, \mathcal{G}_c, s_c, h_c) = 1$ may be a candidate for new wetland development.

IV. Methodology

In this section, we describe our ***Po**tential wetlands via **T**ransfer learning and **A**daptive propagation (PoTA)* algorithm, designed to tackle the data sparsity challenge in wetland prioritization. Figure 2 provides an overview of the overall architecture, where we introduce the key components of the model below:

- **(Feature Processing – Section IV-A)** Input features (for both source and target regions) are vectorized through discretization or normalization based on their types.
- **(Transfer Learning – Section IV-B)** Three latent feature extractors (LFEs) are applied to these features to capture domain-aware knowledge. Specifically, a domain discriminator aids the extraction of two sets of (source and target) domain-specific latent features along with one set of shared latent features. The shared latent features serve as the bridge that enables knowledge transfer from the (wetland-rich) source region to the target region.
- **(Adaptive Propagation – Section IV-C)** In the next step, we apply adaptive propagation on top of the extracted latent features. In particular, the proposed algorithm learns whether information propagation between the two cells should be positive or whether the information impact of one cell on another should be negative.

A. Feature Processing and Extraction

1) Soil Data: This dataset includes discrete classes like flood frequency and elements with continuous values, like available water storage. We encode these features as follows:

- **Discrete features** are one-hot encoded. Note that, some discrete features, like flood frequency, are ordinal rather than categorical. In this implementation, we ignore the distinction between these two types of discrete values and use one-hot encoding for both.
- **Continuous features,** such as water storage, are normalized to the range $[0, 1]$ by dividing to the maximum value among all the cells in the source and target regions.

2) HAND Data: The HAND dataset is binarized using a $2m$ threshold. Specifically, for any cell where the height above the nearest drainage is greater than or equal to $2m$, the value of h_c is set to 1; otherwise, h_c is set to 0.

B. Transfer Learning

In this section, we introduce the transfer learning techniques we developed to effectively bring information learned from the wetland-rich source region to the target region with few wetlands. In particular, the encoded input feature vectors for both source and target regions are fed into the two layers of neural networks, referred to as latent feature extractors
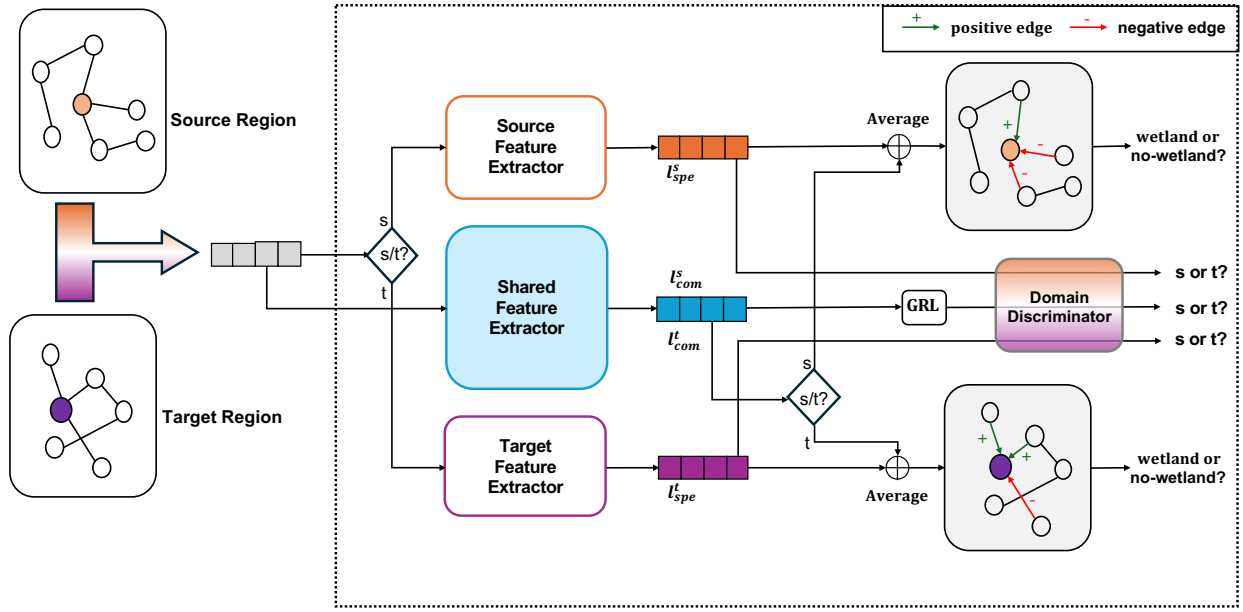
Fig. 2: Overview of the proposed *Po*tential wetlands via *T*ransfer learning and *A*daptive propagation (PoTA)

(LFEs). As shown in Figure 2, these three LFEs generate four outputs; source-domain specific latent features of the source (l_{spe}^s), target-domain specific latent features of the target (l_{spe}^t), shared-domain (or common-domain) latent features of the source (l_{com}^s), and shared-domain latent features of the target (l_{com}^t). In this subsection, we present detailed strategies to guide this process for effectively extracting domain-specific and shared-domain latent features.

Domain Adaptation [32] aims to reduce the differences in latent features extracted from two domains. While such an approach may help find domain shareable features, we also need to identify the unique characteristics of regions for wetlands (e.g., Arizona is extremely dry due to its desert climate) [33]. Many approaches have been proposed recently [16], [17], [19], [34], but in PoTA, we develop a domain disentanglement strategy, as in [11], to capture domain-specific knowledge along with shareable information. To achieve this, a Gradient Reversal Layer (GRL), which applies a negative multiplier on the back-propagated weights during training, is used before the extracted latent features are fed into a domain discriminator to increase the entropy between the learned latent features. The key is that by applying GRL only to shared-domain features, the domain discriminator is trained to distinguish between the specific characteristics of the two domains effectively. PoTA leverages two layers of fully connected neural networks as the domain discriminator. As can be seen in Figure 2, for each cell in the source or target region, two features, l_{com}^* and l_{spe}^* (where $* \in \{s, t\}$) are passed separately as inputs to the domain discriminator:

$$\widehat{d}_{com}^* = F_{disc}(l_{com}^*), \quad \widehat{d}_{spe}^* = F_{disc}(l_{spe}^*). \quad (1)$$

$\widehat{d}^*$ represents the predicted domain probability. Then, the loss

is calculated through binary cross-entropy as,

$$\mathcal{L}_{com}^* = -\frac{1}{N_*} \sum_{s=1}^{N_*} log(1 - \widehat{d}_{com}^*),$$

$$\mathcal{L}_{spe}^* = -\frac{1}{N_*} \sum_{s=1}^{N_*} log(1 - \widehat{d}_{spe}^*), \quad (2)$$

where N_s and N_t are the training batches for the source and target regions respectively (we omit the true label in the above equation that is binary for the source and target domain, $d \in \{0, 1\}$). Given this, we define the overall domain loss as,

$$\mathcal{L}_{dom} = \rho(\mathcal{L}_{com}^s + \mathcal{L}_{spe}^s) + (1 - \rho)(\mathcal{L}_{com}^t + \mathcal{L}_{spe}^t), \quad (3)$$

Here, the ratio $\rho = \frac{N_s}{N_s + N_t}$ of the batch sizes serve as the weight for the training of the discriminator concerning the importance of the source domain [35]. During training, the domain discriminator is updated using Equation 3.

C. Adaptive Propagation

The knowledge transfer between local regions is facilitated by a graph neural network (GNN). Let $\mathcal{G}^* = (\mathcal{V}^*, \mathcal{E}^*, X^*)$ be a graph with $|\mathcal{V}^*| = n$ nodes, $|\mathcal{E}^*| = m$ edges, and feature matrix X^*. As before, when considering the source region, $* = s$, and when considering the target region, we have $* = t$.

The label matrix corresponding to the given region is denoted as $Y^* \in \mathbb{R}^{n \times C}$, where $C(= 2)$ stands for the number of classes. $A^* \in \{0, 1\}^{n \times n}$ denotes the adjacency matrix for the undirected graph $\mathcal{G}^*$, where the degree of node i is represented as $\delta_i = \sum_{j=1}^{n} A_{ij}^*$. The feature matrix, $X^* \in \mathbb{R}^{n \times h}$, where h is the number of dimensions of the latent space. Given the above, the representation of $\mathcal{V}^*$ is updated through message-passing between neighboring nodes.

GNNs have the advantage of addressing data sparsity by incorporating the characteristics of neighboring nodes into their predictions. However, if neighboring cells have very different characteristics, this can lead to performance degradation as the propagated information may serve as noise, rather than enrichment. Therefore, here we propose an adaptive propagation technique to address this issue. This is achieved by varying the edge weights considering the features of the two neighboring cells corresponding to the edge. We first average the domain-specific and shared-domain features for all nodes in the given region (i.e., $v_i \in \mathcal{V}^*$):

$$l_i^* = \frac{l_{spe,i}^* + l_{com,i}^*}{2}, \quad (4)$$

Next, given an edge between $i \leftrightarrow j$, we determine the corresponding edge weight through an attention mechanism:

$$w_{ij}^* = tanh\left((l_i^* + l_j^*)\vec{a}_*\right), \quad (5)$$

where $\vec{a}_*$ is a learnable attention vector; $tanh$ helps ensure that edge weights satisfy the constraint $-1 \leq w_{ij}^* \leq 1$. Given these edge coefficients, following [30], we update node features as follows:

$$l_i^{*,(\gamma)} = l_i^{*,(0)} + \sum_{j \in \mathcal{N}_i} \frac{w_{ij}^*}{\sqrt{\delta_i \delta_j}} l_j^{*,(\gamma-1)}, \quad (6)$$

where, $\mathcal{N}_i$ denotes the set of neighbors of node i with incoming edges into i. As defined earlier, δ and (γ) denote the degree of a node and hidden layer, respectively.

D. Classification

Assuming L layers of propagation, the final output of the adaptive propagation is an enriched latent representation $l_i^{*,(L)}$ for $v_i \in \mathcal{V}^*$. This is fed into a fully connected network, $W^* \in \mathcal{R}^{h \times C}$, with input dimensionality h, and output dimension of $C(=2)$, corresponding to "wetland" and "no-wetland", respectively.

Given the final output, $f^{*,(L)} \in \mathcal{R}^{N_* \times C}$ (where N_s and N_t are the training batches for the source and target regions respectively), the wetland probability is given by:

$$f^{*,(L)} = l^{*,(L)} W^*,$$

where we can define the prediction loss using the negative log-likelihood function as follows:

$$\mathcal{L}_{pred}^* = -\frac{1}{N_*} \sum_{s=1}^{N_*} \sum_{k=1}^{C} y_{sk} \log(f_{sk}^{*,(L)}). \quad (7)$$

E. Optimization and Inference

(Optimization) We optimize the model using domain loss, $\mathcal{L}_{dom}$ (Eq. 3), and prediction loss, $\mathcal{L}_{pred}$ (Eq. 7) as below:

$$\mathcal{L} = \mathcal{L}_{pred}^s + \mathcal{L}_{pred}^t + \lambda \mathcal{L}_{dom}. \quad (8)$$

Here, λ adjusts the weight of domain loss to help stable convergence. During optimization, we employ the Adam optimizer with early stopping based on the validation score.

(Inference) After convergence, we forward the features of each cell in a target domain to the common and target LFEs, followed by the adaptive GNN for wetland prioritization.

F. Theoretical Analysis

This section provides theoretical background on why domain disentanglement and adaptive propagation are necessary for this task.

Theorem 4.1 (Domain disentanglement): Let us assume a domain identifier $\mathcal{D} \in \{\mathcal{S}, \mathcal{T}\}$. Regardless of a specific domain $\mathcal{D}$, the mutual information (I) between the domain-common feature l_{com}^* and domain-specific one l_{spe}^* [36] can be decomposed as below:

$$I(l_{spe}^*; l_{com}^*) = -I(\mathcal{D}; l_{spe}^*, l_{com}^*) + I(\mathcal{D}; l_{spe}^*) + I(\mathcal{D}; l_{com}^*) \quad (9)$$

Then, we get $I(\mathcal{D}; l_{spe}^*, l_{com}^*) = I(\mathcal{D}; l_{spe}^*) + I(\mathcal{D}; l_{com}^*) - I(l_{spe}^*; l_{com}^*)$ with a slight modification. Since $I(\mathcal{D}; l_{com}^*)$ is maximized by a domain discriminator and $I(l_{spe}^*; l_{com}^*)$ is minimized through the objective function (wetland identification), we can infer $I(\mathcal{D}; l_{spe}^*, l_{com}^*) \geq I(\mathcal{D}; l_{spe}^*)$ that is more informative for training.

Theorem 4.2 (Adaptive propagation): Adaptive propagation determines the weight and sign of edges based on features. While a positive signed edge has the effect of smoothing between connected nodes, the negative edges increase the separability. Let us assume two nodes i, j connected with a positive edge, where the label of node i is k. Given the negative likelihood loss $\mathcal{L}_{nll}(Y_i, \widehat{Y}_i)_k = -\log(\widehat{y}_{i,k})$, the gradient of node i is defined as $\partial \mathcal{L}_{nll}(Y_i, \widehat{Y}_i)_k / \partial \widehat{y}_{i,k}$. Similarly, the gradient of neighboring node j follows:

$$\nabla_j \mathcal{L}_{nll}(Y_i, \widehat{Y}_i)_k = \frac{\partial \mathcal{L}_{nll}(Y_i, \widehat{Y}_i)_k}{\partial \widehat{y}_{i,k}} \cdot \frac{\partial \widehat{y}_{i,k}}{\partial h_{\mathbf{i},k}^{(L)}} \cdot \frac{\partial h_{\mathbf{i},k}^{(L)}}{\partial h_{\mathbf{j},k}^{(L)}} \quad (10)$$

Since $\partial h_{\mathbf{i},k}^{(L)} / \partial h_{\mathbf{j},k}^{(L)} > 0$ with positive connection, we can infer that node j will get closer to node i proportional to the $|\eta \partial h_{\mathbf{i},k}^{(L)} / \partial h_{\mathbf{j},k}^{(L)}|$. Vice versa, the gradient has the opposite sign but the same scale.

G. Computational Complexity

Our model consists of two main components; feature extraction with domain disentanglement and adaptive propagation. The first module can be approximated as $\mathcal{O}((A + B) \cdot N \cdot e)$, where A and B refer to the time of forward passing in feature extraction and domain discriminator. N is the input size and e is the number of training epochs. The cost of the second module is dominated by the cost of the GNN, $\mathcal{O}(|\mathcal{E}|\theta_{GNN})$, which is proportional to the number, $|\mathcal{E}|$, of edges in the considered graphs and the number, θ_{GNN}, of trainable parameters. Since our method further employs edge weight retrieval using the coordinates, the complexity becomes $\mathcal{O}(|\mathcal{E}|\theta_{GNN} + |\mathcal{E}|\theta_{Adapt})$. Thus, the cost of the entire module can be $\mathcal{O}((A + B) \cdot N \cdot e + |\mathcal{E}|(\theta_{GNN} + \theta_{Adapt}))$.

V. Experiments

In this section, we investigate three key questions that characterize the proposed method and provide a comprehensive analysis:

TABLE II: (RQ1) Accuracy (Acc. %) and Recall (%) for 3 target regions – bold with underline indicate best Accuracy and recall. A symbol $*$ indicates a significance level (ρ-value< 0.05)

Type	Method	Metric	Texas (sparse) AZ	WA	FL	Oregon (moderate) AZ	WA	FL	Louisiana (dense) AZ	WA	FL
Single-Domain	MLP (2 layers)	Acc.		82.0			85.9			89.3	
		Recall		62.6			94.6			99.0	
	GCN [37]	Acc.		82.8			91.0			91.3	
		Recall		63.1			92.0			99.1	
	GAT [38]	Acc.		80.7			84.7			90.4	
		Recall		68.3			77.8			98.8	
	FAGCN [30]	Acc.		82.5			91.8			91.5	
		Recall		62.9			86.7			99.5	
Cross-Domain	DAREC [14]	Acc.	81.2	81.3	84.0	84.4	86.5	86.2	88.4	89.4	89.4
		Recall	62.0	62.3	68.9	79.3	82.1	96.2	96.6	97.3	96.8
	MMT [39]	Acc.	82.0	83.0	83.1	86.2	86.6	86.0	89.1	91.3	91.3
		Recall	63.7	66.0	67.7	92.3	93.6	90.5	98.5	99.1	99.4
	SER [11]	Acc.	83.5	84.0	84.3	87.2	86.4	86.8	91.0	90.1	89.0
		Recall	66.6	67.2	67.5	91.2	91.4	92.8	99.3	98.3	97.1
	DH-GAT [40]	Acc.	83.4	85.1	86.5	88.8	90.7	90.2	90.6	91.5	91.4
		Recall	70.4	74.0	74.9	95.8	92.0	97.3	99.1	99.7	99.6
Ours	PoTA	Acc.	84.1*	85.4*	**87.2**<u></u>*	92.1*	**93.0**<u></u>*	92.6*	92.2*	92.3*	**92.5**<u></u>*
		Recall	70.8*	74.2*	**77.0**<u></u>*	96.0*	97.8*	**98.1**<u></u>*	99.5*	99.6*	**99.8**<u></u>*

TABLE III: Details of the benchmark datasets

Domain	Region	# wetland cells	# cells
Source	Arizona (AZ, sparse)	6.05 K (0.6 %)	1.05 M
	Washington (WA, moderate)	23.2 K (2.2 %)	1.05 M
	Florida (FL, dense)	0.12 M (12 %)	1.04 M
Target	Texas (TX, sparse)	1.21 K (0.1 %)	1.12 M
	Oregon (OR, moderate)	15.6 K (1.5 %)	1.05 M
	Louisiana (LA, dense)	0.09 M (8.3 %)	1.05 M

TABLE IV: (RQ1) Accuracy gains due to knowledge transfer: gains are higher when the knowledge is transferred from denser regions (LA and FL) to sparser regions (TX and AZ)

		Target TX:0.1%	AZ:0.6%	OR:1.5%	WA:2.2%	LA:8.3%	FL:12%	AVG
Source	TX:0.1%	0.0%	1.1%	0.7%	0.3%	0.4%	0.5%	0.6%
	AZ:0.6%	0.6%	0.0%	0.8%	0.0%	0.3%	0.3%	0.4%
	OR:1.5%	2.1%	2.1%	0.0%	0.6%	0.3%	0.6%	1.1%
	WA:2.2%	1.9%	2.6%	1.6%	0.0%	0.4%	0.5%	1.4%
	LA:8.3%	5.6%	3.1%	0.9%	1.2%	0.0%	0.7%	2.3%
	FL:12%	3.7%	5.2%	1.2%	0.9%	0.6%	0.0%	2.3%
	AVG	2.8%	2.8%	1.0%	0.6%	0.4%	0.5%	

- **RQ1:** Does the proposed model achieve good performance compared to state-of-the-art baselines?
- **RQ2:** How much do domain disentanglement and adaptive propagation contribute to the overall performance? Does the shareable knowledge help to solve the sparsity issue?
- **RQ3:** Does domain disentanglement effectively distinguish between domain-specific and domain-shareable features?

A. Datasets, Baselines, and Setup

(Datasets) In Section III-A, we introduced the details of three datasets; NLCD, Soil, and HAND. In this section, we consider 6 regions of the US with varying wetland characteristics Table III.

(Baselines) We consider 8 state-of-the-art methods, both single-domain and cross-domain, listed in Table II. As a hyperparameter, we set $\lambda = 0.2$ in Eq. 8 for our model. In addition, the learning ratio is set as $1e^{-3}$ with the early stopping of 300 epochs for all baselines.

(Experimental setup) Training/validation/testing sets are split 10%/40%/50% for all methods. We utilized PyTorch, torch-geometric, and a single GPU (Nvidia Titan Xp) for evaluation.

B. Comparison with Baselines (RQ1)

In Table II, we present accuracy and recall results to compare PoTA to the baselines. Firstly, we see that PoTA provides the best accuracy and recall. We see that, as expected, for the sparse target (Texas), cross-domain methods achieve better results, while the performance gap is smaller in dense domains (Oregon and Louisiana). In particular, knowledge transfer from the dense (Florida) to the sparse (Texas) domain significantly improves the quality of prediction. Along with knowledge transfer, we claim that adaptive propagation also plays an important role in performance improvement. This can be observed when using Arizona as the source domain (sparse). Despite our model gaining almost no benefit from knowledge transfer, significant performance improvements can be seen compared to GCN [37] and GAT [38] across three datasets. The high-level overview of the need for knowledge transfer in wetland prioritization can be summarized as follows. These results are confirmed in Table IV: while knowledge transfer always provides positive gains in accuracy, sparser regions benefit most from being the target, whereas denser regions are best used as the source.

TABLE V: (RQ2) Accuracy and recall to show the impact of excluding domain disentanglement (w/o DD) and adaptive propagation (w/o AP) from the original model PoTA

Methods	TX (from FL)		OR (from WA)		LA (from FL)	
	Accuracy	Recall	Accuracy	Recall	Accuracy	Recall
w/o DD	84.4	73.6	88.3	92.1	91.4	99.4
w/o AP	84.9	76.1	87.6	90.5	90.9	99.0
PoTA	87.2	77.0	93.0	97.8	92.5	99.8

(a) feature ablation

(b) propagation ablation
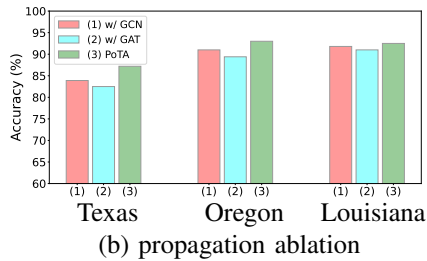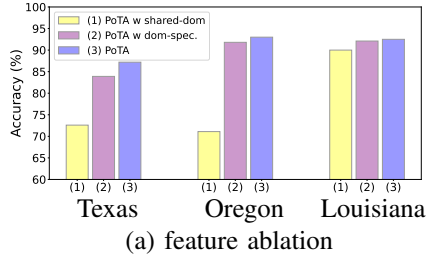
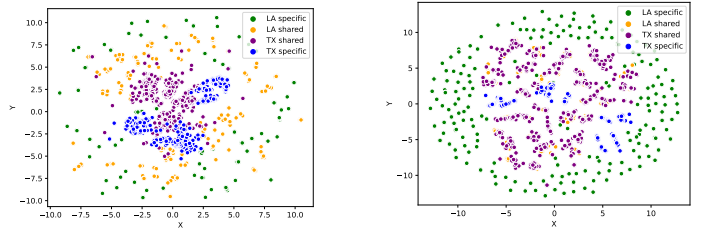Fig. 3: (RQ2) We measure the accuracy to show the impact of (a) using domain-specific vs. domain-shareable features and (b) adaptive propagation

(a) w/o domain disentanglement   (b) w/ domain disentanglement

Fig. 4: (RQ3) t-distributed Stochastic Neighbor Embedding (t-SNE) based visualization of the latent features on Texas (TX). Louisiana (LA) is the source domain

C. Impact of Disentanglement, Sharing, and Adaptive Propagation (RQ2)

One exception in Table II is for the target OR, for which the nearby region, WA, with moderate wetlands, has a greater impact as a source than the wetland-rich region, FL. This makes sense as WA is likely to share more with OR, which can help boost predictions in OR even though it has fewer wetlands than FL. This illustrates the importance of the shared-domain knowledge.

In Table V, we present an ablation study to test the efficacy of domain disentanglement and adaptive propagation. Specifically, we either remove the domain loss in Eq. 3 for PoTA (w/o DD) or propagation in Eq. 6 for PoTA (w/o AP). For each of the three targets, we select the source domain that achieved the best performance in Table II. From the result, we see that excluding domain loss, which hampers region-to-region knowledge transfer, leads to the lowest performance for the sparse domain (Texas). However, it can be found that adaptive propagation plays a more crucial role for denser regions (Oregon and Louisiana) by enabling within-region knowledge transfer.

In Figure 3, we further investigate the impacts of the selected latent features (upper) or the propagation scheme (lower). As seen in the upper chart, as expected, (2) domain-specific training yields better performance than training only

with (1) shared-domain features. However, (3) PoTA effectively combines advantages of the both features for superior performance (Theorem 4.1).

For the lower chart of Figure 3, we replace adaptive propagation with GCN [37] and GAT [38] style propagation: as we see in the chart, adaptive propagation provides the highest accuracy, emphasizing the necessity of distinguishing between the like and unlike neighbors during information exchange (Theorem 4.2).

D. Impact of Feature Disentanglement (RQ3)

To demonstrate the contribution of domain disentanglement to the separation of extracted features, we sampled 500 cells from each domain. Then, we visualized the latent vectors from feature extractors in Figure 4. Specifically, these vectors comprise source-specific (l_{spe}^s), source-shared (l_{shr}^s), target-specific (l_{spe}^t), and target-shared (l_{shr}^t) features. Referring to Figure 4a, it is evident that the shared features (l_{shr}^s and l_{shr}^t) do not properly overlap. Additionally, some source-specific features (l_{spe}^s) overlap with the target-specific ones (l_{spe}^t), indicating a lack of effective disentanglement. Conversely, Figure 4b demonstrates that shared domain features overlap with each other, while domain-specific features are well separated from the domain-shared information.

VI. CONCLUSIONS

Wetlands are essential, yet they are insufficient in many regions. Existing approaches to prioritizing potential future wetland locations rely heavily on expert knowledge. While data-driven techniques show promise to help guide experts, most U.S. wetlands account for only about 6% of the land cover, making data sparsity a major challenge. In this paper, we propose addressing this through region-to-region and within-region knowledge transfer through domain dis-entanglement and adaptive propagation. This led to a 6.1% improvement in accuracy where wetlands are sparse. We believe this could significantly benefit dry regions, like the southwestern US.

REFERENCES

[1] J. Dewitz *et al.*, "National land cover database (nlcd) 2019 products," *US Geological Survey*, vol. 10, p. P9KZCM54, 2021.

[2] J. G. Lyon and L. K. Lyon, "Wetland identification and delineation," *Boca Raton, Florida*, 1993.

[3] D. A. Hammer and R. K. Bastian, "Wetlands ecosystems: natural water purifiers?" in *Constructed wetlands for wastewater treatment*. CRC Press, 2020, pp. 5–19.

[4] R. W. Tiner, "The concept of a hydrophyte for wetland identification," *Bioscience*, vol. 41, no. 4, pp. 236–247, 1991.

[5] J. L. Richardson, *Wetland soils*, 2001.

[6] S. Mahdavi, B. Salehi, J. Granger, M. Amani, B. Brisco, and W. Huang, "Remote sensing for wetland classification: A comprehensive review," *GIScience & remote sensing*, vol. 55, no. 5, pp. 623–658, 2018.

[7] R. W. Tiner, *Wetland indicators: A guide to wetland formation, identification, delineation, classification, and mapping*. CRC press, 2016.

[8] S. Greenhill, H. Druckenmiller, S. Wang, D. A. Keiser, M. Girotto, J. K. Moore, N. Yamaguchi, A. Todeschini, and J. S. Shapiro, "Machine learning predicts which rivers, streams, and wetlands the clean water act regulates," *Science*, vol. 383, no. 6681, pp. 406–412, 2024.

[9] C. B. Do and A. Y. Ng, "Transfer learning for text classification," *Advances in neural information processing systems*, vol. 18, 2005.

[10] D. S. Maitra, U. Bhattacharya, and S. K. Parui, "Cnn based common approach to handwritten character recognition of multiple scripts," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 1021–1025.

[11] Y. Choi, J. Choi, T. Ko, H. Byun, and C.-K. Kim, "Based domain disentanglement without duplicate users or contexts for cross-domain recommendation," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 293–303.

[12] C. Zhao, C. Li, R. Xiao, H. Deng, and A. Sun, "Catn: Cross-domain recommendation for cold-start users via aspect transfer network," *arXiv preprint arXiv:2005.10549*, 2020.

[13] F. Zhu, Y. Wang, C. Chen, J. Zhou, L. Li, and G. Liu, "Cross-domain recommendation: challenges, progress, and prospects," *arXiv preprint arXiv:2103.01696*, 2021.

[14] F. Yuan, L. Yao, and B. Benatallah, "Darec: deep domain adaptation for cross-domain recommendation via transferring rating patterns," *arXiv preprint arXiv:1905.10760*, 2019.

[15] H. Bonab, M. Aliannejadi, A. Vardasbi, E. Kanoulas, and J. Allan, "Cross-market product recommendation," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 110–119.

[16] Z. Li, Y. Tang, W. Li, and Y. He, "Learning disentangled representation with pairwise independence," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 4245–4252.

[17] P. Nema, A. Karatzoglou, and F. Radlinski, "Disentangling preference representations for recommendation critiquing with ß-vae," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 1356–1365.

[18] J. Cao, X. Lin, X. Cong, J. Ya, T. Liu, and B. Wang, "Disencdr: Learning disentangled representations for cross-domain recommendation," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022, pp. 267–277.

[19] Y. Choi, "Based cross-domain recommendation via hyperbolic embedding and hierarchy-aware domain disentanglement," *arXiv preprint arXiv:2403.20298*, 2024.

[20] M. Neumann, R. Garnett, C. Bauckhage, and K. Kersting, "Propagation kernels: efficient graph kernels from propagated information," *Machine learning*, vol. 102, pp. 209–245, 2016.

[21] A. Garcia Duran and M. Niepert, "Learning graph representations with embedding propagation," *Advances in neural information processing systems*, vol. 30, 2017.

[22] H. Pei, B. Wei, K. C.-C. Chang, Y. Lei, and B. Yang, "Geom-gcn: Geometric graph convolutional networks," *arXiv preprint arXiv:2002.05287*, 2020.

[23] Y. Choi, J. Choi, T. Ko, H. Byun, and C.-K. Kim, "Finding heterophilic neighbors via confidence-based subgraph matching for semi-supervised node classification," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 283–292.

[24] J. Zhu, Y. Yan, L. Zhao, M. Heimann, L. Akoglu, and D. Koutra, "Beyond homophily in graph neural networks: Current limitations and effective designs," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7793–7804, 2020.

[25] J. Zhu, R. A. Rossi, A. Rao, T. Mai, N. Lipka, N. K. Ahmed, and D. Koutra, "Graph neural networks with heterophily," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 11 168–11 176.

[26] X. Li, R. Zhu, Y. Cheng, C. Shan, S. Luo, D. Li, and W. Qian, "Finding global homophily in graph neural networks when meeting heterophily," *arXiv preprint arXiv:2205.07308*, 2022.

[27] Y. Sun, H. Deng, Y. Yang, C. Wang, J. Xu, R. Huang, L. Cao, Y. Wang, and L. Chen, "Beyond homophily: Structure-aware path aggregation graph neural network," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, L. D. Raedt, Ed. International Joint Conferences on Artificial Intelligence Organization, 7 2022, pp. 2233–2240, main Track. [Online]. Available: https://doi.org/10.24963/ijcai.2022/310

[28] T. Wang, D. Jin, R. Wang, D. He, and Y. Huang, "Powerful graph convolutional networks with adaptive propagation mechanism for homophily and heterophily," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 2022, pp. 4210–4218.

[29] X. Zheng, M. Zhang, C. Chen, Q. Zhang, C. Zhou, and S. Pan, "Autoheg: Automated graph neural network on heterophilic graphs," *arXiv preprint arXiv:2302.12357*, 2023.

[30] D. Bo, X. Wang, C. Shi, and H. Shen, "Beyond low-frequency information in graph convolutional networks," *arXiv preprint arXiv:2101.00797*, 2021.

[31] Y. Guo and Z. Wei, "Clenshaw graph neural networks," *arXiv preprint arXiv:2210.16508*, 2022.

[32] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[33] X. Chen, S. Wang, M. Long, and J. Wang, "Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation," in *International conference on machine learning*. PMLR, 2019, pp. 1081–1090.

[34] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf, "Measuring statistical dependence with hilbert-schmidt norms," in *International conference on algorithmic learning theory*. Springer, 2005, pp. 63–77.

[35] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine learning*, vol. 79, no. 1, pp. 151–175, 2010.

[36] H. Hwang, G.-H. Kim, S. Hong, and K.-E. Kim, "Variational interaction information maximization for cross-domain disentanglement," *Advances in Neural Information Processing Systems*, vol. 33, pp. 22 479–22 491, 2020.

[37] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[38] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *stat*, vol. 1050, p. 20, 2017.

[39] A. Krishnan, M. Das, M. Bendre, H. Yang, and H. Sundaram, "Transfer learning via contextual invariants for one-to-many cross-domain recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 1081–1090.

[40] J. Xu and Y. Cai, "Decoupled hyperbolic graph attention network for cross-domain named entity recognition," in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 591–600.