Non-IID Transfer Learning on Graphs

Jun Wu[1], Jingrui He[1], Elizabeth Ainsworth[1,2]

[1]University of Illinois at Urbana-Champaign, [2]USDA ARS Global Change and Photosynthesis Research Unit

junwu3@illinois.edu, jingrui@illinois.edu, ainswort@illinois.edu

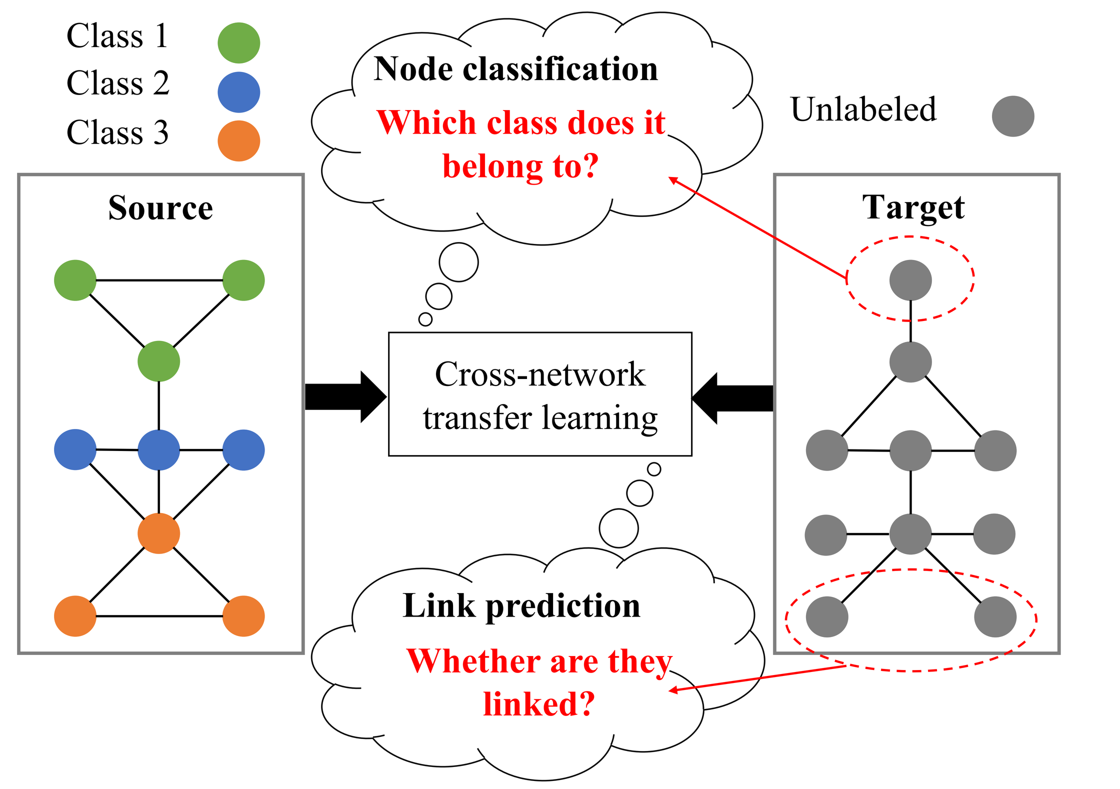
AIFARMS

Artificial Intelligence for Future Agricultural Resilience, Management, and Sustainability

Problem Definition

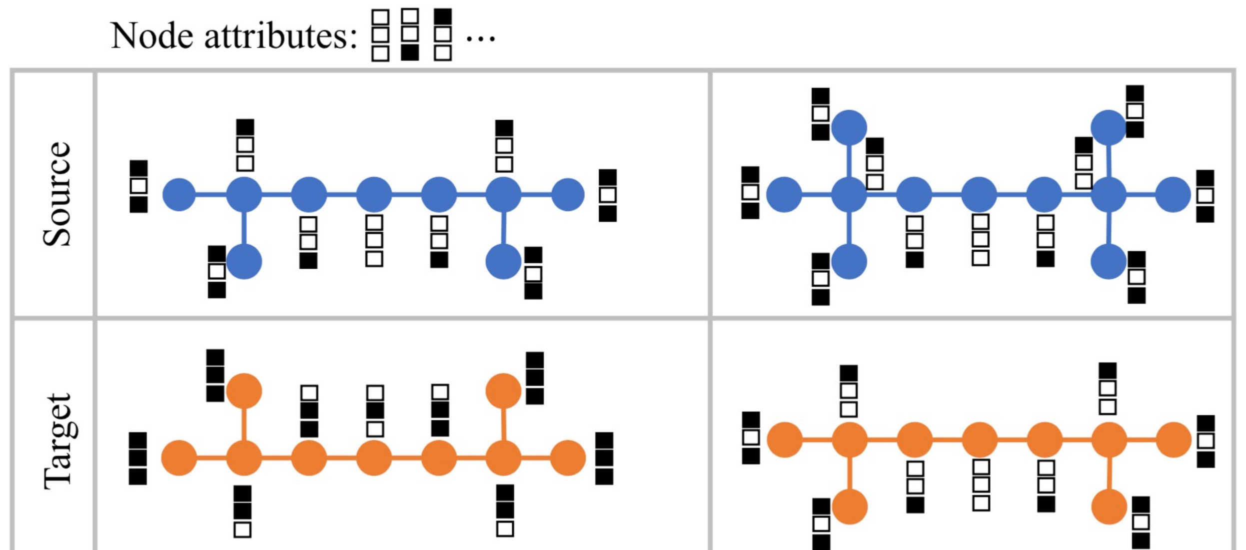
❑ Cross-Network Transfer Learning

- Input: (i) source graph $G^s = (V^s, E^s, X^s)$ with rich structure or node label information; (ii) target graph $G^t = (V^t, E^t, X^t)$ with sparse structure or unlabeled nodes
- Goal: **Learn a prediction function on the target graph**



❑ Graph Distribution Shift (Challenges)

Node attributes:

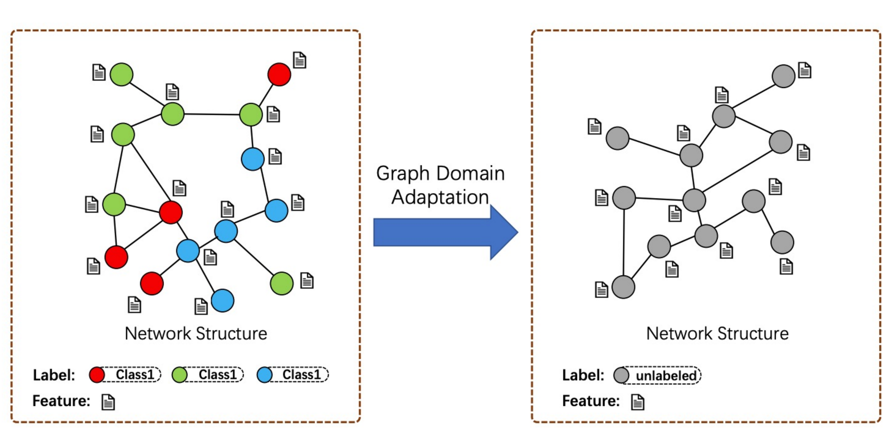


(a) Node attribute distribution shift (b) Graph structure shift

❑ Applications

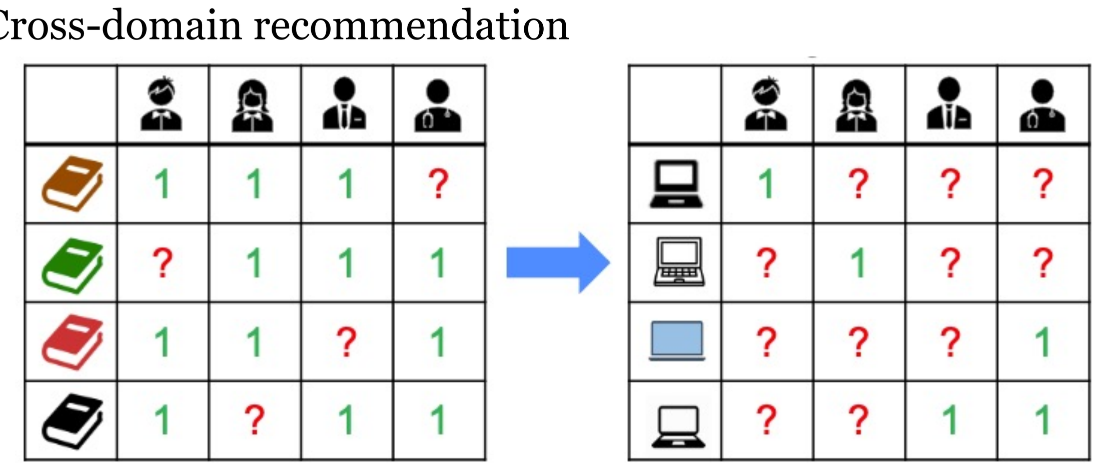
- Cross-network node classification

Source



Target

- Cross-domain recommendation

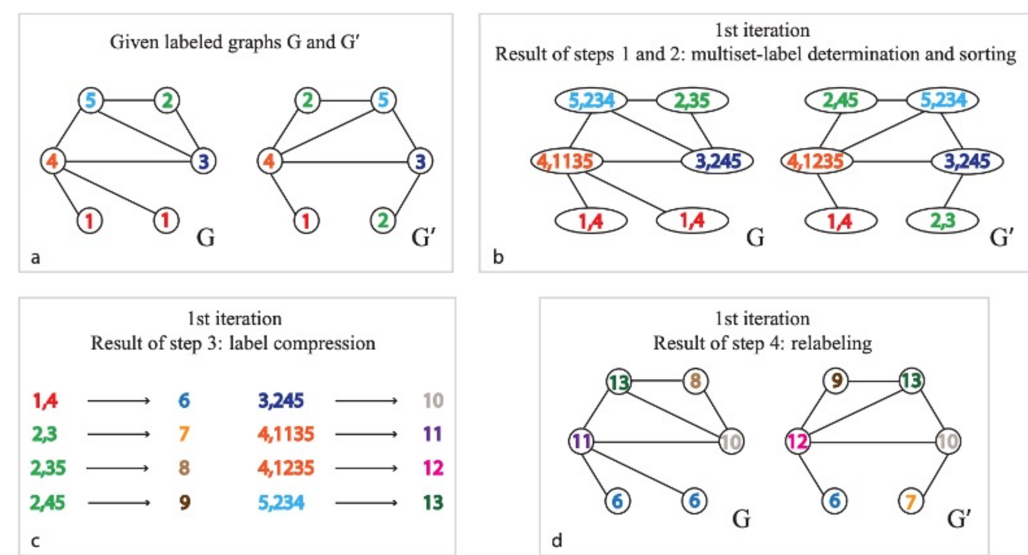


Source **Target**

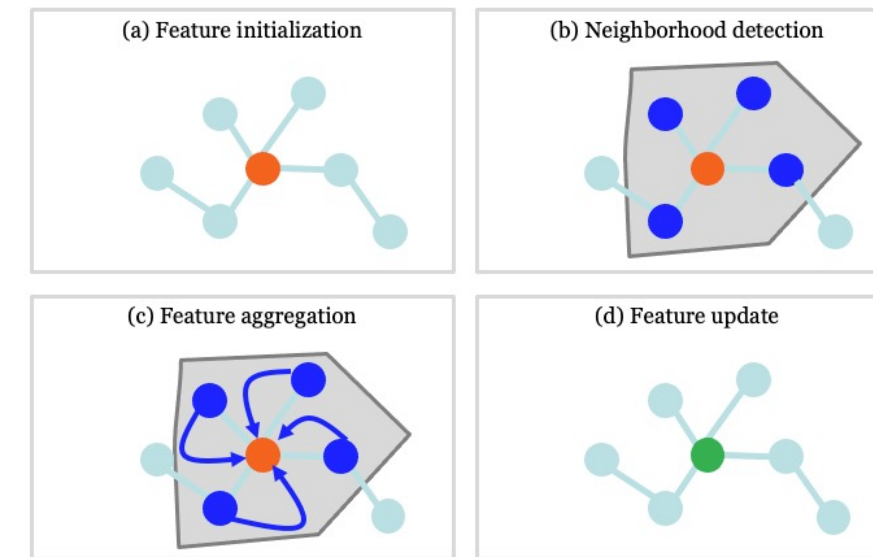
Graph Subtree Discrepancy

❑ Motivation

- Connection between Weisfeiler-Lehman graph kernels [1] and graph neural networks [2]



Weisfeiler-Lehman Graph Subtree Kernel



Message-Passing Graph Neural Network

❑ Formal Definition of Graph Subtree Discrepancy

$$d_{GSD}(G^s, G^t) = \lim_{M\to\infty} \frac{1}{M+1} \sum_{m=0}^{M} d_b(G_m^s, G_m^t)$$

- $d_b(\cdot,\cdot)$: Base distribution discrepancy, e.g., Maximum Mean Discrepancy (MMD) [3]
- G_m^s (G_m^t): Source (target) Weisfeiler-Lehman subgraph at depth m

❑ Generalization Analysis

[*Cross-Network Node Classification*] Given graph subtree discrepancy $d_{GSD}(G^s, G^t)$ defined above, a message-passing GNN with the feature extractor f and the hypothesis $h \in \mathcal{H}$, the node classification error in the target graph can be bounded:

$$\epsilon_t(h \circ f) \le \epsilon_s(h \circ f) + d_{GSD}(G^s, G^t) + \lambda^* + R^*$$

where λ^* measure the labeling difference across domains and R^* is the Bayes error.

[*Cross-Network Link Prediction*] The loss of link prediction is defined as $\epsilon^{link}(h \circ f) = \mathbb{E}_{(u,v)\in V\times V}[L(h(f(u)||f(v)), y)]$ where $y = 1$ if u and v are linked, $y = 0$ otherwise. Then, the link prediction error in the target graph can be bounded:

$$\epsilon_t^{link}(h \circ f) \le \epsilon_s^{link}(h \circ f) + d_{GSD}(G^s, G^t) + \lambda_{link}^* + R_{link}^*$$

where λ_{link}^* measure the labeling difference across domains and R_{link}^* is the Bayes error.

Proposed Algorithm: GRADE

❑ Objective function

$$\min_\theta \underbrace{C(G^s, G^t; \theta)}_{\text{Task-specific Loss}} + \lambda \cdot \underbrace{d_{GSD}(G^s, G^t; \theta)}_{\text{Distribution Discrepancy}}$$

Task-specific Loss **Distribution Discrepancy**

- Cross-network node classification (GRADE-N)

$$\min_{\theta_f, \theta_h} \frac{1}{|V^s|} \sum_{v \in V^s} L\big(f(u^s; \theta_f), y^s; \theta_h\big) + \lambda \cdot d_{GSD}\big(f(G^s; \theta_f), f(G^t; \theta_f)\big)$$

Evaluation

❑ Performance Evaluation

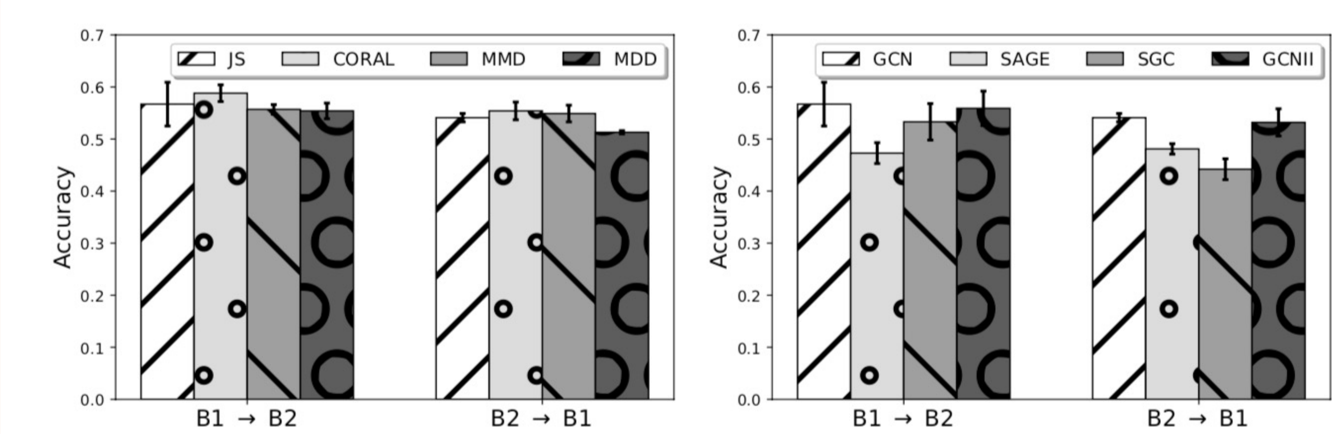
Methods	USA → Brazil	USA → Europe	Brazil → USA	Brazil → Europe	Europe → USA	Europe → Brazil	Avg.
GCN	$0.366_{\pm 0.011}$	$0.371_{\pm 0.004}$	$0.491_{\pm 0.011}$	$0.452_{\pm 0.012}$	$0.439_{\pm 0.001}$	$0.298_{\pm 0.022}$	0.403
SGC	$0.527_{\pm 0.022}$	$0.430_{\pm 0.009}$	$0.432_{\pm 0.005}$	$0.479_{\pm 0.000}$	$0.447_{\pm 0.002}$	$0.481_{\pm 0.011}$	0.466
GCNII	$0.344_{\pm 0.086}$	$0.393_{\pm 0.025}$	$0.470_{\pm 0.056}$	$0.494_{\pm 0.018}$	$0.460_{\pm 0.012}$	$0.542_{\pm 0.011}$	0.450
DAN	$0.504_{\pm 0.020}$	$0.393_{\pm 0.000}$	$0.436_{\pm 0.006}$	$0.393_{\pm 0.010}$	$0.436_{\pm 0.003}$	$0.542_{\pm 0.000}$	0.451
DANN	$0.500_{\pm 0.005}$	$0.386_{\pm 0.011}$	$0.402_{\pm 0.048}$	$0.350_{\pm 0.062}$	$0.436_{\pm 0.000}$	$0.538_{\pm 0.005}$	0.435
MDD	$0.500_{\pm 0.005}$	$0.378_{\pm 0.000}$	$0.402_{\pm 0.048}$	$0.350_{\pm 0.062}$	$0.402_{\pm 0.048}$	$0.477_{\pm 0.081}$	0.418
AdaGCN	$0.466_{\pm 0.065}$	$0.434_{\pm 0.004}$	**$0.501_{\pm 0.003}$**	$0.486_{\pm 0.021}$	$0.456_{\pm 0.034}$	$0.561_{\pm 0.081}$	0.484
UDA-GCN	**$0.607_{\pm 0.059}$**	$0.388_{\pm 0.007}$	$0.497_{\pm 0.005}$	**$0.510_{\pm 0.019}$**	$0.434_{\pm 0.042}$	$0.477_{\pm 0.024}$	0.486
EGI	$0.523_{\pm 0.013}$	$0.451_{\pm 0.011}$	$0.417_{\pm 0.021}$	$0.454_{\pm 0.046}$	$0.452_{\pm 0.029}$	**$0.588_{\pm 0.011}$**	0.481
GRADE-N	$0.550_{\pm 0.062}$	**$0.457_{\pm 0.027}$**	$0.497_{\pm 0.010}$	$0.506_{\pm 0.004}$	**$0.463_{\pm 0.001}$**	**$0.588_{\pm 0.032}$**	**0.510**

Cross-network node classification on airport data set

Methods	CD → Music			Music → CD			Book → Movie			Movie → Book		
	HR@10	MRR@10	NDCG@10	HR@10	MRR@10	NDCG@10	HR@10	MRR@10	NDCG@10	HR@10	MRR@10	NDCG@10
BPRMF	$0.182_{\pm 0.003}$	$0.061_{\pm 0.003}$	$0.089_{\pm 0.003}$	$0.259_{\pm 0.008}$	$0.097_{\pm 0.006}$	$0.134_{\pm 0.007}$	$0.198_{\pm 0.003}$	$0.070_{\pm 0.002}$	$0.099_{\pm 0.002}$	$0.128_{\pm 0.007}$	$0.040_{\pm 0.003}$	$0.060_{\pm 0.004}$
NeuMF	$0.286_{\pm 0.012}$	$0.104_{\pm 0.007}$	$0.145_{\pm 0.008}$	$0.328_{\pm 0.014}$	$0.109_{\pm 0.015}$	$0.160_{\pm 0.013}$	$0.294_{\pm 0.020}$	$0.102_{\pm 0.006}$	$0.146_{\pm 0.009}$	$0.142_{\pm 0.009}$	$0.045_{\pm 0.006}$	$0.066_{\pm 0.007}$
CoNet	$0.405_{\pm 0.018}$	$0.161_{\pm 0.002}$	$0.214_{\pm 0.006}$	$0.333_{\pm 0.010}$	$0.119_{\pm 0.018}$	$0.168_{\pm 0.015}$	$0.319_{\pm 0.023}$	$0.116_{\pm 0.020}$	$0.162_{\pm 0.020}$	$0.142_{\pm 0.025}$	$0.041_{\pm 0.011}$	$0.064_{\pm 0.014}$
CGN	$0.357_{\pm 0.018}$	$0.120_{\pm 0.015}$	$0.175_{\pm 0.015}$	$0.476_{\pm 0.042}$	$0.192_{\pm 0.020}$	$0.255_{\pm 0.026}$	$0.359_{\pm 0.032}$	$0.136_{\pm 0.018}$	$0.187_{\pm 0.021}$	$0.205_{\pm 0.014}$	$0.071_{\pm 0.004}$	$0.102_{\pm 0.006}$
PPGN	$0.419_{\pm 0.016}$	$0.179_{\pm 0.008}$	$0.231_{\pm 0.009}$	$0.564_{\pm 0.044}$	$0.278_{\pm 0.032}$	$0.336_{\pm 0.035}$	$0.489_{\pm 0.011}$	$0.239_{\pm 0.007}$	$0.291_{\pm 0.007}$	$0.296_{\pm 0.007}$	$0.154_{\pm 0.010}$	$0.184_{\pm 0.009}$
EGI	$0.446_{\pm 0.011}$	$0.196_{\pm 0.003}$	$0.250_{\pm 0.006}$	$0.599_{\pm 0.007}$	$0.265_{\pm 0.015}$	$0.338_{\pm 0.009}$	$0.461_{\pm 0.021}$	$0.224_{\pm 0.017}$	$0.274_{\pm 0.014}$	$0.274_{\pm 0.024}$	$0.147_{\pm 0.008}$	$0.174_{\pm 0.011}$
GRADE-R	**$0.450_{\pm 0.006}$**	**$0.197_{\pm 0.002}$**	**$0.251_{\pm 0.003}$**	**$0.600_{\pm 0.011}$**	**$0.313_{\pm 0.007}$**	**$0.373_{\pm 0.008}$**	**$0.505_{\pm 0.022}$**	**$0.249_{\pm 0.004}$**	**$0.302_{\pm 0.008}$**	**$0.318_{\pm 0.006}$**	**$0.170_{\pm 0.009}$**	**$0.202_{\pm 0.006}$**

Cross-domain recommendation on Amazon data set

❑ Impact of Base Distribution Discrepancy and Base GNN



Conclusion

- ❑ **Problem**: Cross-network transfer learning where knowledge is transferred from a source graph to relevant a target graph
- ❑ **Algorithm**: Graph adaptive network (GRADE) which measures the graph distribution shift using graph subtree structures
- ❑ **Evaluation**: Superior performance on both cross-network node classification and cross-domain recommendation

References

[1] Nino Shervashidze, Pascal Schweitzer, Erik Jan Van Leeuwen, Kurt Mehlhorn, and Karsten M. Borgwardt. "Weisfeiler-lehman graph kernels." *Journal of Machine Learning Research*, 2011.

[2] Thomas N. Kipf, and Max Welling. "Semi-supervised classification with graph convolutional networks." *ICLR*, 2017.

[3] Gretton, Arthur, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. "A kernel two-sample test." *Journal of Machine Learning Research* , 2012.

ILLINOIS