Matcha: Mitigating Graph Structure Shifts with Test-Time Adaptation

Wenxuan Bao [1] Zhichen Zeng [1] Zhining Liu [1] Hanghang Tong [1] Jingrui He [1]

[1]University of Illinois Urbana-Champaign

Problem Formulation

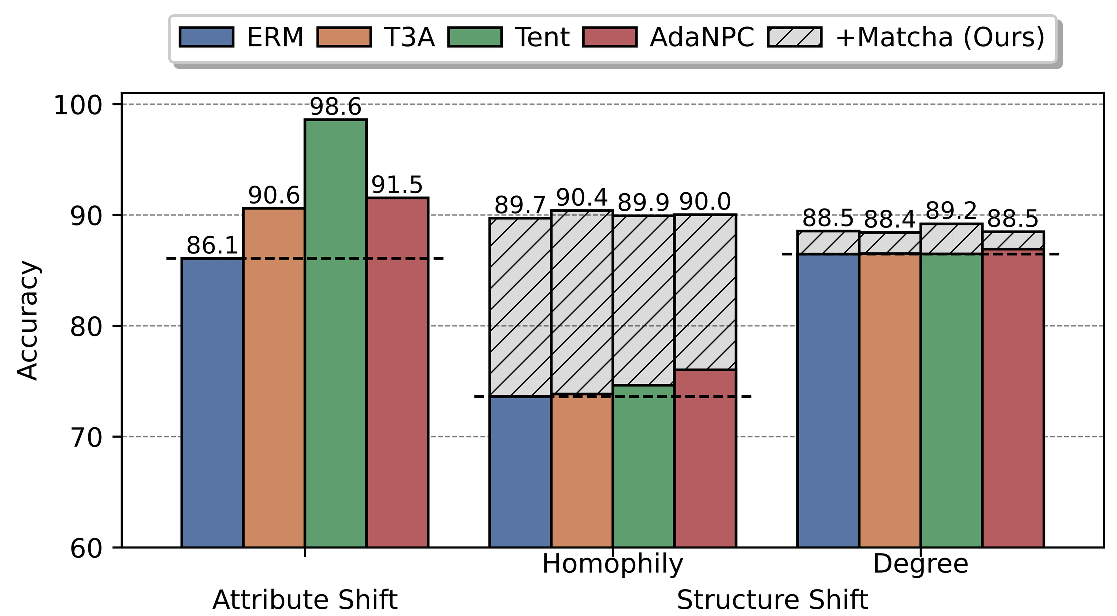
Graph neural networks (GNNS) are vulnerable to **distribution shifts**, including

- **(Node) Attribute Shifts**: Node feature distributions are different.
- **(Graph) Structure Shifts**: Node connectivity patterns are different.

Structure shifts include, but are not limited to

- *Degree Shifts*: Changes in average node degree.
- *Homophily Shifts*: Changes in average node homophily.

Test-Time Adaptation (TTA) addresses distribution shifts by adapting a source model to the target domain, without access to source data.



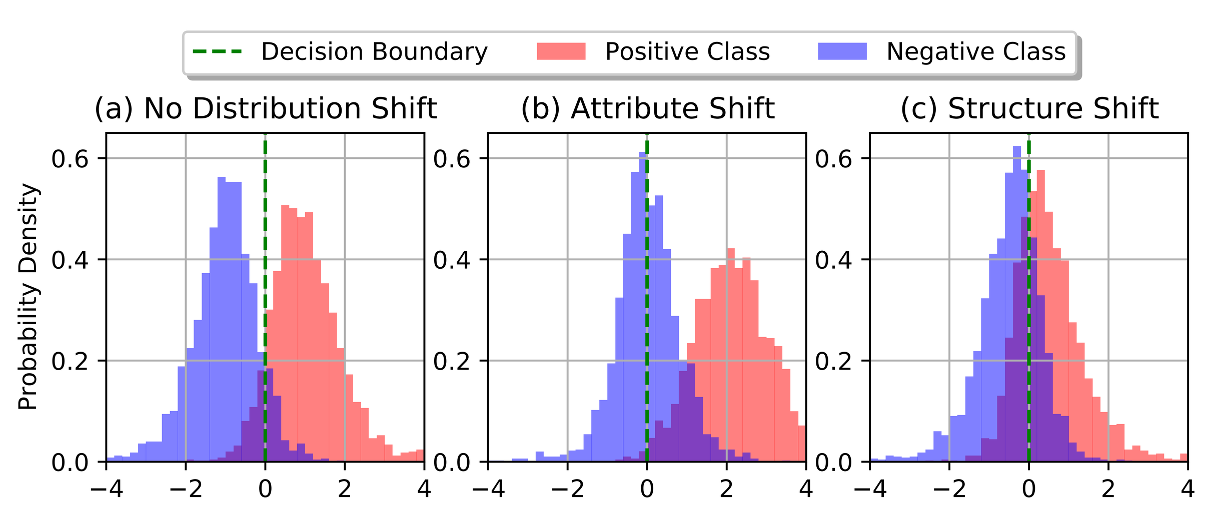
Many existing generic TTA methods (T3A, Tent, AdaNPC) are developed for images.

- These methods perform well under attributes shifts.
- However, they often fail under structure shifts.

Key Questions: Why does this performance gap exist? How can we enhance the performance of TTA under structure shifts?

Observation

We visualize the distribution of node representations before the last linear classification layer, after projected to 1D.



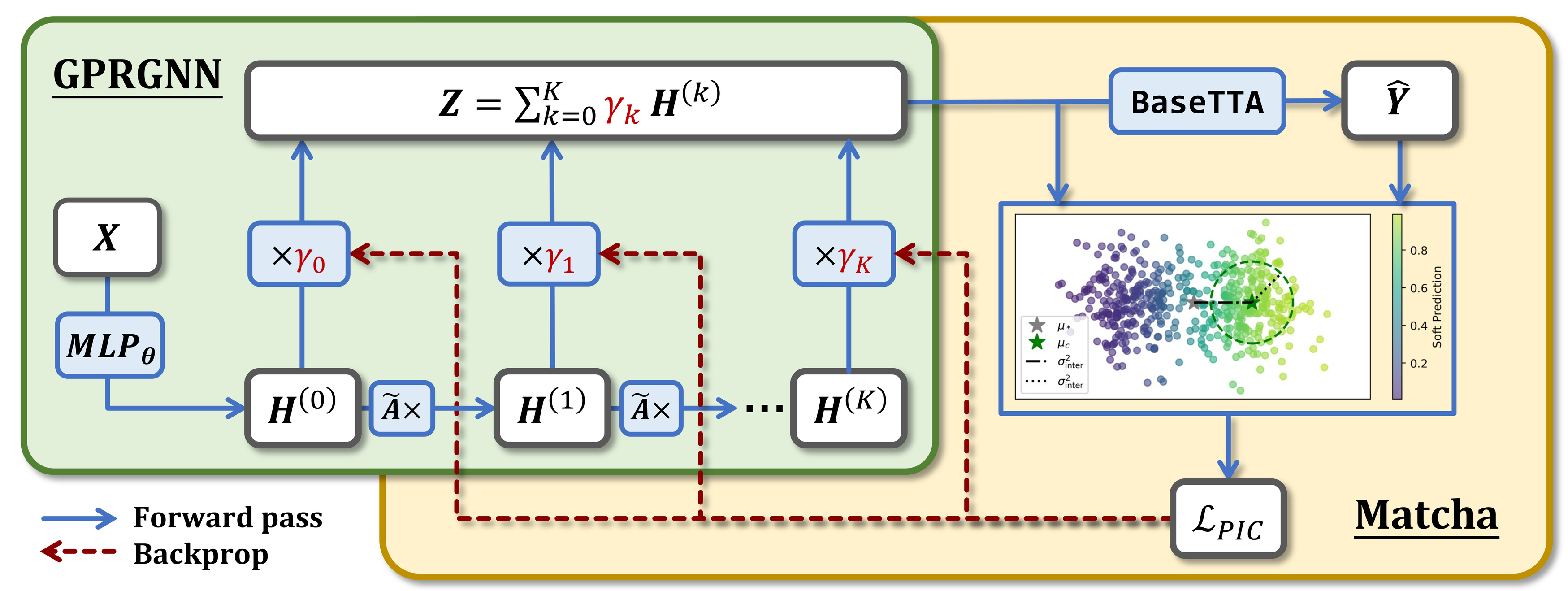
Histograms of Projected Representations Under Different Distribution Shifts

Key observation: Attribute shifts and structure shifts have different impact patterns!

- Attribute shifts mainly introduce **classifier bias**. Although the distribution of node representations changes, making the original decision boundary suboptimal, the quality of the representations remains intact. Adjusting the decision boundary can recover the original accuracy.
- Structure shifts mainly introduce **representation degradation**. They result in increased overlap between node representations from different classes, reducing class separability. As a result, adjusting the decision boundary alone is insufficient to recover accuracy, explaining the poor performance of generic TTA under structure shifts.

Our theoretical analysis verifies these observations with single-layer GCN and CSBM graphs.

Proposed Method: Matcha



GPRGNN [1] aggregates the 0 to K-hop representation $\boldsymbol{H}^{(0)}, \cdots, \boldsymbol{H}^{(K)}$ with parameters γ. Structure shifts does not affect $\boldsymbol{H}^{(0)}$, but change the signal-to-noise ratio in $\boldsymbol{H}^{(1)}, \cdots, \boldsymbol{H}^{(K)}$. **We treat γ as the *hop-aggregation parameters* to be updated.**

Prediction-Informed Clustering Loss encourages nodes of the same class to have similar representations and push apart those of different classes. Given a test graph with C classes, M nodes and their representations $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_M$, and soft predictions $\{\hat{y}_{i,c}\}_{i,c}$:

$$\mathcal{L}_{\text{PIC}} = \frac{\sigma^2_{\text{intra}}}{\sigma^2_{\text{intra}} + \sigma^2_{\text{inter}}}, \quad \text{where}$$

- $\sigma^2_{\text{intra}} = \sum_{i=1}^{M} \sum_{c=1}^{C} \hat{y}_{i,c} \|\boldsymbol{z}_i - \boldsymbol{\mu}_c\|_2^2$ is the *intra-class variance*, measuring the variance of node representations within each (pseudo) class.
- $\sigma^2_{\text{inter}} = \sum_{c=1}^{C} \left(\sum_{i=1}^{M} \hat{y}_{i,c}\right) \|\boldsymbol{\mu}_c - \boldsymbol{\mu}_*\|_2^2$ is the *inter-class variance*, measuring how far different class centroids are from the global centroid.
- $\boldsymbol{\mu}_c = \frac{\sum_{i=1}^{M} \hat{y}_{i,c} \boldsymbol{z}_i}{\sum_{i=1}^{M} \hat{y}_{i,c}}$ is the centroid of class c, $\boldsymbol{\mu}_* = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{z}_i$ is the centroid for all nodes.

Compatibility to Generic TTA In each optimization step, Matcha

- First apply a base TTA algorithm to get prediction $\hat{\boldsymbol{Y}} = \{\hat{y}_{i,c}\}$.
- Then compute the PIC loss with $\hat{\boldsymbol{Y}}$ to optimize node representation.

By alternately updating node representations and predictions, Matcha achieves a synergy between representation quality and prediction accuracy, effectively handling both structure and attribute shifts.

Experiments

CSBM with Controlled Distribution Shifts Matcha consistently enhances the performance of different base TTA methods, across different types of distribution shifts.

Table 1:Accuracy (mean ± s.d. %) on CSBM with structure shifts and attribute shifts. hom - homophilious, het - heterophilious, high - high degree, low - low degree

Method	Homophily shift		Degree shift		Attribute + homophily shift		Attribute + degree shift	
	hom → het	het → hom	high → low	low → high	hom → het	het → hom	high → low	low → high
ERM	73.62 ± 0.44	76.72 ± 0.89	86.47 ± 0.38	92.92 ± 0.43	61.06 ± 1.67	72.61 ± 0.38	77.63 ± 1.13	73.60 ± 3.53
+ Matcha	89.71 ± 0.27	90.68 ± 0.26	88.55 ± 0.44	93.78 ± 0.74	85.34 ± 4.68	74.70 ± 0.99	78.29 ± 1.41	73.86 ± 4.20
T3A	73.85 ± 0.24	76.68 ± 1.08	86.52 ± 0.44	92.94 ± 0.37	65.77 ± 2.11	72.92 ± 0.90	80.89 ± 1.28	81.94 ± 3.24
+ Matcha	90.40 ± 0.11	90.50 ± 0.24	88.42 ± 0.60	93.83 ± 0.41	88.49 ± 0.58	79.34 ± 1.85	81.82 ± 1.36	82.12 ± 4.03
Tent	74.64 ± 0.38	79.40 ± 0.57	86.49 ± 0.50	92.84 ± 0.18	74.42 ± 0.41	79.57 ± 0.40	86.05 ± 0.33	93.06 ± 0.24
+ Matcha	89.93 ± 0.16	**91.26 ± 0.08**	**89.20 ± 0.20**	**94.88 ± 0.09**	**90.12 ± 0.07**	**91.15 ± 0.20**	**87.76 ± 0.16**	**95.04 ± 0.06**
AdaNPC	76.03 ± 0.46	81.66 ± 0.17	86.92 ± 0.38	91.15 ± 0.39	63.96 ± 1.31	76.33 ± 0.71	77.69 ± 0.91	76.24 ± 3.06
+ Matcha	90.03 ± 0.33	90.36 ± 0.67	88.49 ± 0.31	92.84 ± 0.57	85.81 ± 0.30	77.63 ± 1.55	78.41 ± 1.03	76.31 ± 3.68
GTrans	74.01 ± 0.44	77.28 ± 0.56	86.58 ± 0.11	92.74 ± 0.13	71.60 ± 0.60	74.45 ± 0.42	83.21 ± 0.25	89.40 ± 0.62
+ Matcha	89.47 ± 0.20	90.31 ± 0.31	87.88 ± 0.77	93.23 ± 0.52	88.88 ± 0.38	76.87 ± 0.66	83.41 ± 0.16	89.98 ± 0.93
SOGA	74.33 ± 0.18	83.99 ± 0.35	86.69 ± 0.37	93.06 ± 0.21	70.45 ± 1.71	76.41 ± 0.79	81.31 ± 1.03	88.32 ± 1.94
+ Matcha	89.92 ± 0.26	90.69 ± 0.27	88.83 ± 0.32	94.49 ± 0.23	88.92 ± 0.28	90.14 ± 0.33	87.11 ± 0.28	93.38 ± 1.06
GraphPatcher	79.14 ± 0.62	82.14 ± 1.11	87.87 ± 0.18	93.64 ± 0.45	64.16 ± 3.49	76.98 ± 1.04	76.99 ± 1.43	73.31 ± 4.48
+ Matcha	**91.28 ± 0.28**	90.66 ± 0.15	88.01 ± 0.18	93.88 ± 0.69	89.99 ± 0.41	87.94 ± 0.39	78.43 ± 1.84	77.86 ± 4.14

Real-World Experiments We also test Matcha under structure shifts in Syn-Cora, Syn-Products, Twitch-E, and OGB-Arxiv.

Visualization While structure shifts blur the boundary between the node representations of two classes (b), Matcha successfully restores the quality of node representations (f), and outperforms other surrogate losses (c)(d)(e).

Acknowledgement. This work is supported by National Science Foundation under Award No. IIS-2416070, and the U.S. Department of Homeland Security under Grant Award Number 17STQAC00001-08-00. The views and conclusions are those of the authors and should not be interpreted as representing the official policies of the funding agencies or the government.

[1] Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. In *ICLR*, 2021.