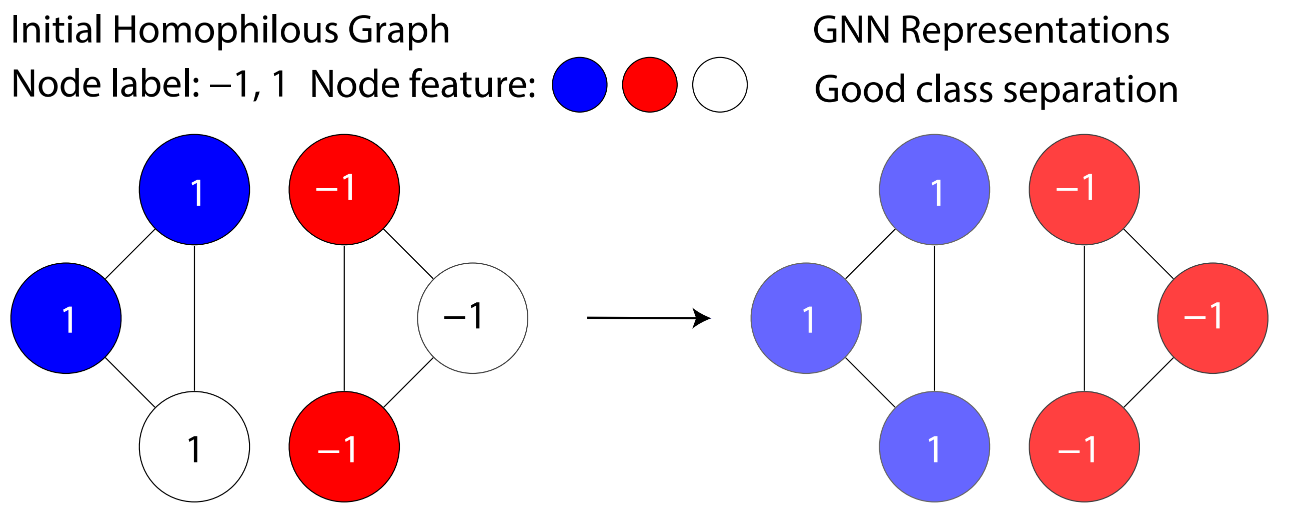
Understanding GNNs and Homophily in Dynamic Node Classification

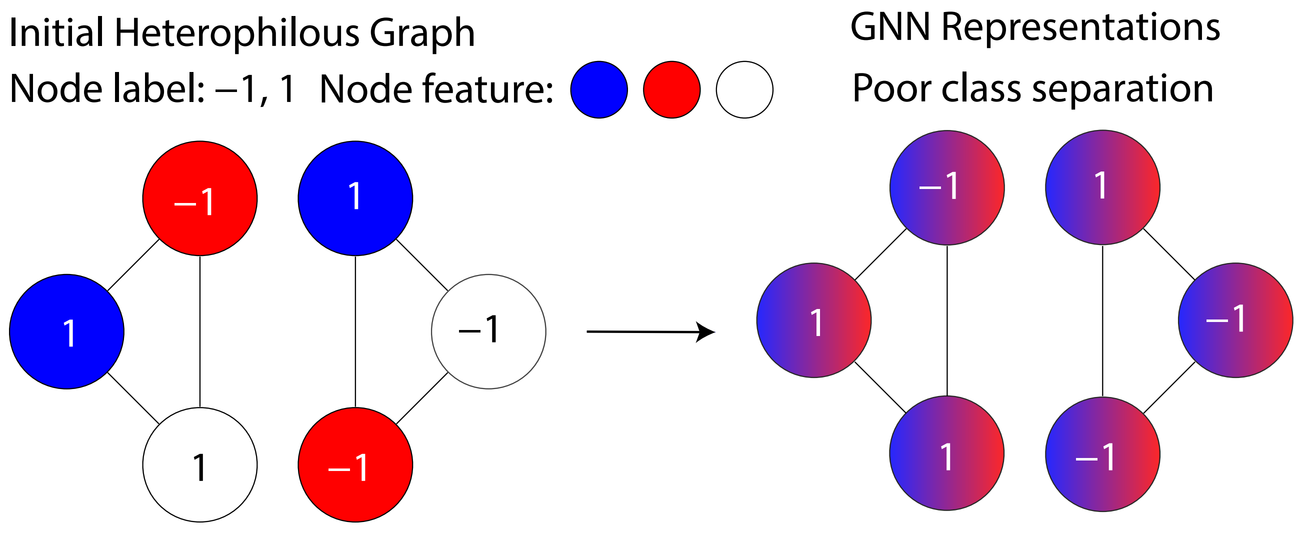
Michael Ito, Danai Koutra, Jenna Wiens
University of Michigan, Ann Arbor, MI, USA
Division of Computer Science and Engineering
contact: mbito@umich.edu

Paper | Code

On highly homophilous graphs → GNNs obtain strong performance



On many heterophilous graphs → GNN performance degrades



$$\text{Static Homophily: } h^S = \mathbb{P}(y(i) = y(j) \mid j \in \mathcal{N}(i))$$

$$\text{GCN Layer } l \;: \mathbf{z}^{(l+1)}(i) = \textsc{Aggregate}^{(l)}(\{\mathbf{z}^{(l)}(j) : j \in \hat{\mathcal{N}}(i)\})$$

Homophily has been crucial in advancing the study of GNNs. However, current analyses are **limited to static graphs**. We analyze homophily in **dynamic settings.**

Theoretical Framework

Assumptions: binary labels $y_t(i) \in \{-1, +1\}$, gaussian features $\mathbf{z_t^0(i)} \in \mathcal{N}(y_t(i) \cdot \mu, \sigma)$, linear GCN.

The expected AUROC from t to $t+1$: $1 - \Phi\left(-\frac{\mathbb{E}_{i|i \in V^+_{t+1}}\left[\mathbf{z}_t^{(\ell)}(i)\right] - \mathbb{E}_{j|j \in V^-_{t+1}}\left[\mathbf{z}_t^{(\ell)}(j)\right]}{\mathbb{V}_{i,j|i \in V^+_{t+1}, j \in V^-_{t+1}}\left[\mathbf{z}_t^{(\ell)}(i) + \mathbf{z}_t^{(\ell)}(j)\right]}\right)$

Characterizing Class Separation

Thm. The expected class separation at time t after l layers of a GCN can be expressed as

$$\mathbb{E}_{i|i \in V^+_{t+1}}[\mathbf{z}_t^{(\ell)}(i)] - \mathbb{E}_{j|j \in V^-_{t+1}}[\mathbf{z}_t^{(\ell)}(j)] = 2 \cdot \mu_t \cdot (h_t^+ + h_t^- - 1)^\ell$$

where $h_t^+ = \mathbb{P}(y_{t+1}(i) = y_t(j) \mid j \in \hat{\mathcal{N}}_t(i), y_{t+1}(i) = +1)$,

$h_t^- = \mathbb{P}(y_{t+1}(i) = y_t(j) \mid j \in \hat{\mathcal{N}}_t(i), y_{t+1}(i) = -1)$

Insight: class separation decays exponentially in l at the rate $h_t^+ + h_t^- - 1 \rightarrow$ *oversmoothing* occurs at the rate of homophily levels

Characterizing Class Variances

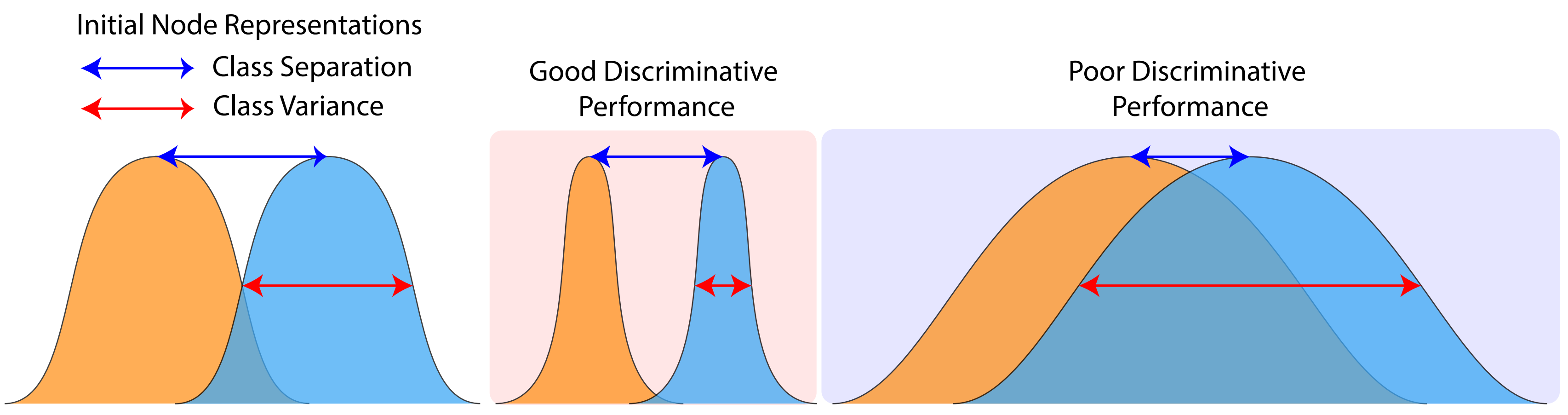
Thm. The lower bounds of class variances at time t after l layers of a GCN can be expressed as

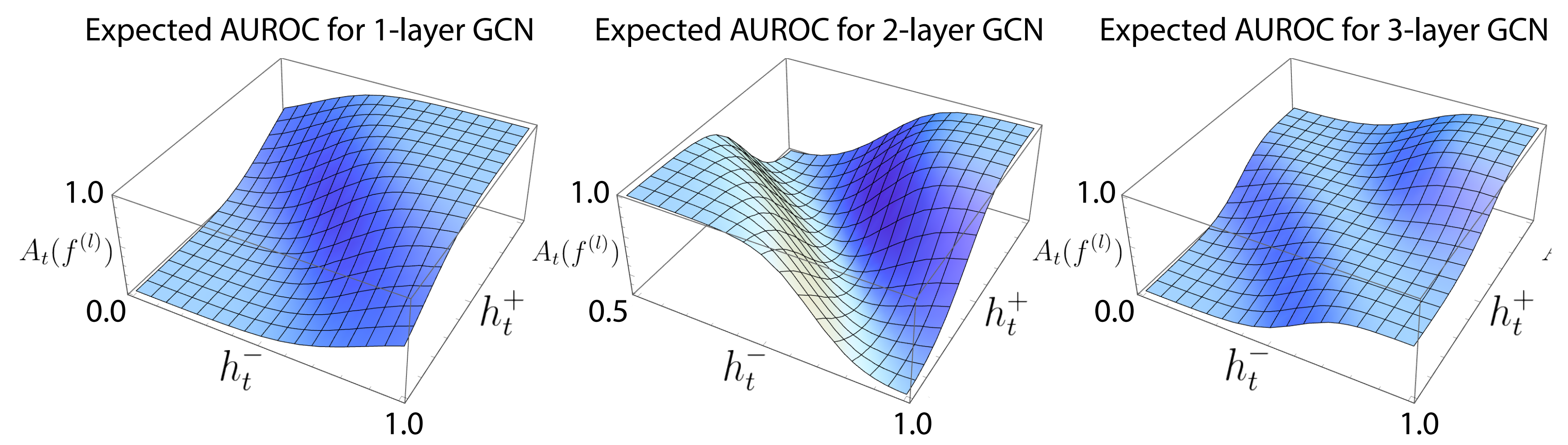
$$\mathbb{V}_{i|i \in V^+_{t+1}}\left[\mathbf{z}_t^{(\ell)}(i)\right] = {h_t^+}^2 \cdot \mathbb{V}_{i|i \in V^+_{t+1}}\left[\mathbf{z}_t^{(\ell-1)}(i)\right] + (1 - h_t^+)^2 \cdot \mathbb{V}_{j|j \in V^-_{t+1}}\left[\mathbf{z}_t^{(\ell-1)}(j)\right]$$

$$\mathbb{V}_{j|j \in V^-_{t+1}}\left[\mathbf{z}_t^{(\ell)}(j)\right] = {h_t^-}^2 \cdot \mathbb{V}_{j|j \in V^-_{t+1}}\left[\mathbf{z}_t^{(\ell-1)}(j)\right] + (1 - h_t^-)^2 \cdot \mathbb{V}_{i|i \in V^+_{t+1}}\left[\mathbf{z}_t^{(\ell-1)}(i)\right]$$

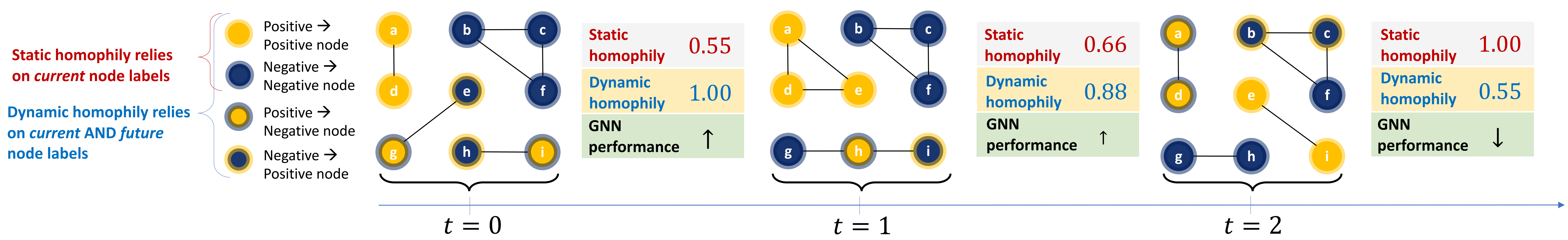
$$\mathbb{V}_{i|i \in V^+_{t+1}}\left[\mathbf{z}_t^{(0)}(i)\right] = \mathbb{V}_{j|j \in V^-_{t+1}}\left[\mathbf{z}_t^{(0)}(j)\right] = \sigma_t^2.$$

Insight: Variance lower bounds allow us to upper bound AUC → the best GNN AUC for different homophily levels are achieved by different GNN layers



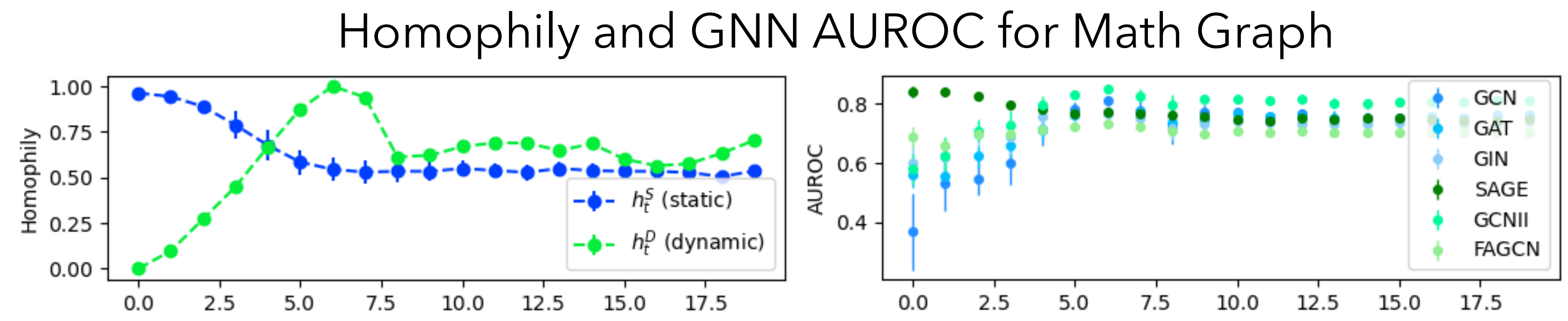


New Homophily Metric for Dynamic Graphs: $h_t^D = \frac{1}{|C|}\sum_{c \in C} \mathbb{P}(y_{t+1}(i) = y_t(j) \mid j \in \hat{\mathcal{N}}_t(i), y_{t+1}(i) = c)$

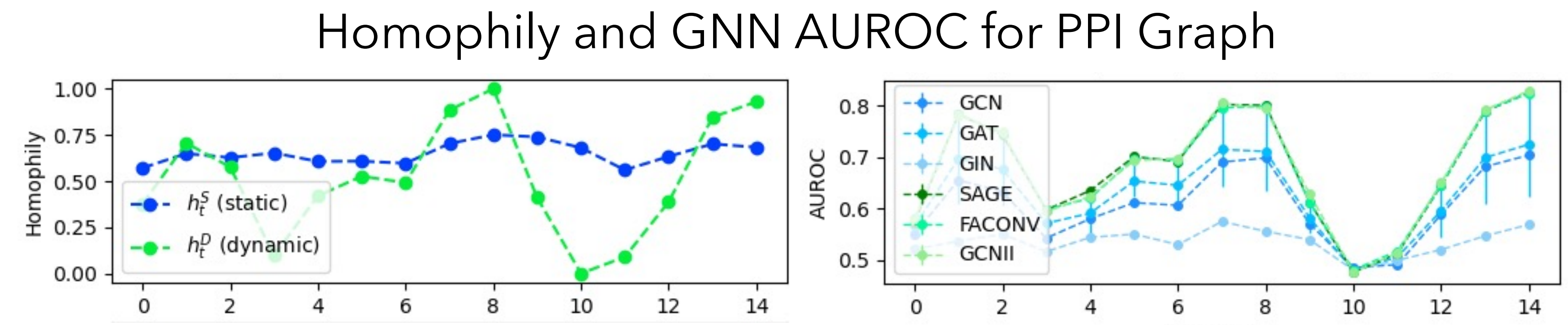


Dynamic homophily accurately correlates with GNN discriminative performance

Homophily and GNN AUROC for Math Graph



Homophily and GNN AUROC for PPI Graph



We show popular GNNs **struggle under low dynamic homophily**, motivating new GNNs for dynamic graphs. Our work further lays a foundation for future dynamic GNNs by **characterizing GNN discriminative power in dynamic settings**

Funding: This work was supported by a graduate fellowship from the Department of Energy (DE-SC0023112)