

Popularity-driven random walk graph convolutional networks

Narendran R. V 

*School of Physics, Indian Institute of Science Education and Research Thiruvananthapuram
(IISER TVM), Thiruvananthapuram 695551, Kerala, India*

Alwin Poulose *

*School of Data Science, Indian Institute of Science Education and Research Thiruvananthapuram
(IISER TVM), Thiruvananthapuram 695551, Kerala, India*



(Received 19 May 2026; accepted 2 September 2026; published 24 September 2026)

We introduce the popularity-driven random walk graph convolutional network (PDRW-GCN), a topology-aware graph neural network that incorporates degree-biased diffusion to exploit the heavy-tailed connectivity of scale-free networks. By incorporating a structural bias parameter, β , the model dynamically reweights neighborhood propagation based on node degree. We identify a topology-dependent diffusion transition in heterophilic graph learning: Sparse heterophilic networks benefit from hub-amplified transport ($\beta > 0$), whereas dense heterophilic networks require hub-suppressed diffusion ($\beta < 0$) to prevent noisy neighborhood overmixing. Experiments across homophilic and heterophilic benchmark networks demonstrate consistent improvements in node classification performance measured by F1 score and receiver operating characteristic—area under the curve. Moreover, the model exhibits diminishing performance gains on non-scale-free topologies, indicating that β acts primarily as a topology-dependent routing mechanism rather than a generic regularizer.

DOI: [10.1103/qgbq-9r7p](https://doi.org/10.1103/qgbq-9r7p)

I. INTRODUCTION

Graph neural networks (GNNs), particularly graph convolutional networks (GCNs), have emerged as a powerful framework for representation learning on graph-structured systems [1,2]. Standard GCNs operate through iterative neighborhood aggregation, implicitly assuming homophily: Connected nodes are likely to share similar features and class labels [3]. While this assumption holds in citation and social networks, performance degrades severely in heterophilic graphs, where neighboring nodes often belong to different classes [4,5].

This limitation becomes particularly pronounced in scale-free networks characterized by heavy-tailed degree distributions and highly connected hub nodes [6,7]. Because hubs participate in a large fraction of local message-passing operations, they strongly influence the global latent representation. In homophilic settings, these hubs act as structural anchors that stabilize information propagation. Under heterophily, however, hubs instead aggregate conflicting neighborhood signals, leading to structural noise and oversmoothing [8,9]. Here, overmixing refers to the aggregation

*Contact author: alwinpoulosepalatty@iisertvm.ac.in

of semantically inconsistent neighborhood information under heterophily, whereas oversmoothing denotes the progressive homogenization of node representations across repeated propagation layers.

Existing approaches to heterophily typically introduce structural corrections, frequency filtering, adaptive PageRank weighting, or learned routing mechanisms [4,10–14]. Graph attention networks (GATs) [10] dynamically weight neighborhood contributions using feature-dependent attention coefficients, whereas Geom-GCN [11] constructs latent geometric neighborhoods to better separate structurally dissimilar nodes. CenGCN [12] explicitly regulates hub influence through node centrality measures. Similarly, H2GCN [4] decouples ego-node representations from higher-order neighborhood aggregations to mitigate neighborhood pollution. From a spectral and structural perspective, frequency adaptation graph convolutional networks (FAGCN) [13] adaptively combines low- and high-frequency signals to accommodate nonsmooth feature spaces, while GPR-GNN [14] optimizes learnable generalized PageRank weights to dynamically balance multihop context. Although these approaches improve performance in certain heterophilic settings, they often introduce significant computational overhead, require multihop feature concatenation, or rely on heuristic architectural adjustments, thereby providing limited physical insight into how underlying topology governs information flow.

A complementary viewpoint is to regard message passing itself as a diffusion process occurring on the underlying graph. Such a physical interpretation naturally connects graph representation learning with the well-established theory of random walks on complex networks, where information propagation is governed directly by network topology. Random walks have long served as a fundamental framework for describing diffusion and transport on complex networks. Recent work on random walk graph neural networks demonstrated that differentiable random walk operations can be incorporated directly into graph learning architectures [15]. While these models demonstrate that random walk operators can be successfully integrated into graph learning architectures, they typically employ unbiased or generic diffusion mechanisms. This raises a natural question: Can physically motivated random walk dynamics provide additional control over information propagation, particularly in heterogeneous networks?

Degree-biased random walks have been extensively investigated in the statistical physics and complex networks literature as a mechanism for controlling transport efficiency, search dynamics, network exploration, and routing behavior [16,17]. Earlier studies demonstrated that modifying transition probabilities based on node degree can substantially alter stationary distributions, mean first-passage times, transport efficiency, and diffusion localization in heterogeneous networks [18,19]. More recent work has further generalized these concepts by introducing nonlinear, popularity-driven diffusion models, providing a physically interpretable framework for topology-dependent transport on scale-free networks [20,21]. These studies, however, primarily focus on characterizing transport processes themselves rather than incorporating degree-biased diffusion directly into graph representation learning. Moreover, Santini [22] and subsequent generalized degree-biased transport studies [23] also showed that diffusion dynamics on scale-free networks can be strongly altered by degree-dependent random walk dynamics, where transition probabilities are explicitly biased by network topology. Collectively, these results suggest that degree bias provides a principled mechanism for regulating diffusion according to network topology. Despite this rich theoretical foundation, such topology-dependent transport mechanisms have received comparatively little attention within graph neural networks, where message passing remains largely architecture driven rather than transport driven.

Motivated by this gap, we reinterpret graph message passing as a topology-dependent degree-biased diffusion process on scale-free networks. We propose the popularity-driven random walk graph convolutional network or PDRW-GCN, which integrates degree-biased diffusion directly into the graph convolutional operation, unlike previous degree-biased random walk studies, which primarily investigate transport characteristics or routing efficiency. By introducing a structural bias parameter β , the proposed framework dynamically regulates hub accessibility during neighborhood propagation, providing a physically interpretable mechanism for topology-aware information transport in graph neural networks. We show that positive bias values improve propagation stability

in sparse heterophilic networks by reinforcing dominant structural pathways. In contrast, negative bias values suppress excessive hub influence and reduce overmixing in dense heterophilic graphs. Furthermore, the model systematically degrades on non-scale-free networks, suggesting that the proposed mechanism primarily operates as a topology-dependent routing process rather than a generic regularization strategy.

The remainder of this paper is organized as follows: Sec. II presents the mathematical formulation of the proposed PDRW-GCN architecture, including the popularity-driven transition matrix, the adaptive self-loop mechanism, the forward-propagation dynamics, and a computational complexity analysis. Section III describes the experimental setup, benchmark datasets, preprocessing pipeline, and evaluation methodology. Section IV presents empirical results across homophilic, heterophilic, and non-scale-free graph regimes, together with ablation studies and computational complexity analysis. Finally, Sec. V summarizes the primary conclusions of this work and outlines several directions for future research.

II. MODEL AND THEORETICAL FRAMEWORK

The proposed PDRW-GCN reformulates graph message passing as a topology-dependent degree-biased diffusion process. We first review the standard GCN formulation before deriving the popularity-driven transition operator and the resulting graph convolutional propagation rule.

A. Standard GCNs

Let $G = (V, E)$ denote an undirected graph with $N = |V|$ nodes, adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, and node-feature matrix $\mathbf{X} \in \mathbb{R}^{N \times F}$, where F is the feature dimension. Standard GCNs perform neighborhood aggregation by symmetrically normalizing the adjacency matrix [1]. The layerwise propagation rule is given by

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)}), \quad (1)$$

where the tilde notation denotes quantities computed after the inclusion of self-loops and $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{A}_{ij}$ is the corresponding degree matrix. From a transport perspective, the normalized propagation operator $\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2}$ may be interpreted as one step of an unbiased or isotropic diffusion process on the graph, where feature information propagates uniformly across neighboring nodes. Consequently, standard GCNs implicitly assume that adjacent nodes carry mutually consistent semantic information. Under heterophily, aggregating features across structurally dissimilar neighborhoods introduces conflicting semantic information. Repeated applications of the propagation operator therefore increase overmixing and eventually lead to oversmoothing, causing node embeddings to become progressively indistinguishable.

B. Popularity-driven random walk

We begin with a standard unbiased random walk, where the probability of moving from node i to any neighbor j is distributed uniformly across all outgoing edges

$$P_{ij}^{(\text{std})} = \frac{A_{ij}}{k_i}, \quad (2)$$

where $k_i = \sum_j A_{ij}$ denotes the degree of node i . This uniform transition process forms the basis of conventional message-passing architectures.

To introduce controlled topological bias, we instead employ a popularity-driven random walk (PDRW), where transition probabilities are scaled according to the degree of the destination node,

$$P_{ij}^{(\text{pdrw})} = \frac{A_{ij} k_j^\beta}{\sum_{m \in \mathcal{N}(i)} A_{im} k_m^\beta}, \quad (3)$$

where $\beta \in \mathbb{R}$ is a structural bias parameter and $\mathcal{N}(i)$ denotes the neighborhood of node i .

The parameter β directly controls the transport dynamics on the graph topology. When $\beta = 0$, the system reduces to an unbiased random walk. A positive β increases the transition probability toward highly connected hubs, effectively localizing diffusion trajectories within dominant structural regions. Conversely, a negative β suppresses hub accessibility and redistributes transport toward lower-degree peripheral neighborhoods.

C. Popularity-driven transition operator

We now reformulate neighborhood aggregation as a degree-biased random walk. Let $\Theta \in \mathbb{R}^{N \times N}$ denote the unnormalized routing matrix whose entries are defined by

$$\Theta_{ij} = \begin{cases} A_{ij} (k_j)^\beta, & k_j > 0, \\ 0, & k_j = 0, \end{cases} \quad (4)$$

where isolated nodes ($k_j = 0$) are assigned zero transition weight, preventing undefined degree exponentiation when $\beta < 0$.

Introducing the diagonal matrix $\mathbf{K}^\beta$ with entries $(\mathbf{K}^\beta)_{ii} = (k_i)^\beta$, the routing matrix can be written compactly as

$$\Theta = \mathbf{A} \mathbf{K}^\beta. \quad (5)$$

To obtain a valid stochastic transition operator, Θ is row normalized. Defining the diagonal normalization matrix $\mathbf{D}_\Theta$ such that

$$(\mathbf{D}_\Theta)_{ii} = \sum_{j=1}^N \Theta_{ij}, \quad (6)$$

the PDRW transition matrix is now formally defined as

$$\mathbf{P} = \mathbf{D}_\Theta^{-1} \Theta. \quad (7)$$

The resulting operator dynamically reshapes information flow according to the underlying topology. From a diffusion perspective, the operator $\mathbf{P}$ acts as a topology-dependent transport kernel whose stationary dynamics are governed by the degree-biased accessibility of structural hubs. Positive values of β increase diffusion localization around highly connected nodes, whereas negative values promote broader exploration of peripheral regions.

From a spectral perspective, the structural bias parameter β modifies the effective diffusion geometry of the graph by altering the accessibility of high-degree nodes during propagation. Positive values of β increase diffusion localization around dominant hubs, whereas negative values redistribute transport toward lower-degree peripheral regions. Consequently, the proposed operator induces a topology-dependent modification of the graph transport dynamics.

D. Self-loops

Under heterophily, neighborhood aggregation can overwhelm a node’s intrinsic feature representation, particularly when large structural hubs are present. To ensure that nodes retain their identity under heterophilic conditions, we implemented a dual-strategy adaptive self-loop mechanism tailored to the dataset’s specific density.

For small, sparse datasets, localized statistics exhibit high variance due to limited neighborhood structure. Thus, we used the global average of the nonzero transition weights to scale the self-loops, ensuring structural stability across the sparse graph. For dense, scale-free networks, global averaging is heavily skewed by extreme hubs. For robustness against this severe degree of skewness, we utilized a localized median aggregation that helps prevent the self-loop from being disproportionately inflated by individual hubs (see Algorithm 1).

In both regimes, the resulting self-loop values are amplified by a constant scaling factor of 2.0, selected empirically to preserve ego features while avoiding excessive self-dominance during

ALGORITHM 1. PDRW-GCN forward pass with adaptive self-loops.

Require: Adjacency matrix A , feature matrix X , bias β , graph type (sparse/dense)

Ensure: Node classification probabilities $Z \in \mathbb{R}^{N \times C}$

Step 1: Graph preprocessing and degree calculation

0: $k \leftarrow \sum_{j=1}^N A_{i,j}$ {Compute degrees, assigning $k = 1$ to isolated nodes}

0: $\Theta \leftarrow A \odot (k^\beta)^T$ {Scale columns by target node degree bias}

Step 2: Dual-strategy self-loops (heterophily correction)

0: **if** graph type is sparse **then**

0: $\mu_{\text{global}} \leftarrow \text{Mean}(\Theta[\Theta > 0])$ {Global average of valid edges}

0: **for** $i = 1$ to N **do**

0: $\Theta[i, i] \leftarrow \mu_{\text{global}} \times 2.0$ {Apply stable global self-loop}

0: **end for**

0: **else if** graph type is dense **then**

0: **for** $i = 1$ to N **do**

0: $m_i \leftarrow \text{Median}(\Theta[i, :][\Theta[i, :] > 0])$ {Robust local median}

0: $\Theta[i, i] \leftarrow m_i \times 2.0$ {Apply skew-resistant local self-loop}

0: **end for**

0: **end if**

Step 3: Row normalization

0: $D_\Theta \leftarrow \sum_{j=1}^N \Theta[i, j]$

0: $P \leftarrow D_\Theta^{-1} \Theta$ {Normalize to yield final transition matrix $\mathbf{P}$ }

Step 4: Neural network forward pass

0: $H^{(1)} \leftarrow \text{Dropout}(\text{ReLU}(P \cdot X \cdot W^{(0)}), p = 0.5)$

0: $Z \leftarrow \text{Softmax}(H^{(1)})$

0: **return** $Z = 0$

heterophilic propagation. The joint influence of the self-loop scaling factor along with β on model performance is examined later in Table V.

E. PDRW-GCN propagation rule

Using the stochastic matrix $\mathbf{P}$, the layerwise propagation dynamics of the PDRW-GCN are defined as

$$\mathbf{H}^{(l+1)} = \sigma(\mathbf{P}\mathbf{H}^{(l)}\mathbf{W}^{(l)}), \quad (8)$$

where $\mathbf{H}^{(l)} \in \mathbb{R}^{N \times d_l}$ denotes the node representation matrix at layer l , with $\mathbf{H}^{(0)} = \mathbf{X}$, and $\mathbf{W}^{(l)}$ is the trainable weight matrix. The operator $\mathbf{P}\mathbf{H}^{(l)}$ constitutes the topological routing phase of the model, dynamically aggregating neighborhood information based on the graph's hub structure.

F. Algorithmic implementation

The complete implementation of the proposed PDRW-GCN framework, including the construction of the popularity-driven transition operator, the adaptive self-loop mechanism, row normalization, and the forward-propagation procedure, is summarized in Algorithm 1.

Although the adaptive self-loop strategy remains empirically motivated, the mechanism was designed to preserve ego-feature stability under heterophilic diffusion while minimizing excessive amplification caused by extreme hub dominance.

G. Complexity analysis

Our experimental results suggest that PDRW-GCN maintains the linear time complexity of a standard GCN, avoiding the bottlenecks associated with graph attention networks (GATs).

For a graph $G = (V, E)$ with input feature dimension d and output dimension d' , the dominant cost of graph convolution arises from the dense feature transformation and sparse neighborhood

aggregation. Standard GCN propagation therefore scales as

$$O(|V|dd' + |E|d'). \quad (9)$$

GATs additionally compute attention coefficients for every edge during each forward pass, introducing substantial per-edge computational and memory overhead on dense graphs [10].

In contrast, the PDRW transition operator depends only on static degree-based edge reweighting. When β is treated as a fixed hyperparameter, the transition matrix $\mathbf{P}$ is computed once during preprocessing. The subsequent propagation step therefore reduces to Eq. (8), yielding the same asymptotic complexity as conventional GCNs while avoiding the additional overhead associated with attention-based routing. Empirical run-time measurements validating this complexity behavior are reported later in Table VII.

III. EXPERIMENTAL SETUP

To evaluate the routing behavior of the proposed PDRW-GCN architecture across different graph regimes, we conducted experiments on various benchmark datasets spanning homophilic, sparse heterophilic, dense heterophilic, and non-scale-free topologies. This section outlines the datasets, preprocessing pipeline, training configuration, and evaluation methodology used throughout the study.

A. Baseline models

To evaluate the efficacy of the proposed PDRW-GCN framework, we benchmark its performance against seven state-of-the-art architectures representing four distinct design paradigms: standard isotropic smoothing (GCN [1], GAT [10]), centrality-driven regulation (CenGCN [12]), spectral and adaptive frequency filtering (FAGCN [13], GPR-GNN [14]), and higher-order structural decoupling (Geom-GCN [11], H2GCN [4]). This diverse baseline suite allows us to systematically evaluate degree-biased diffusion against competing homophilic and heterophily aware mechanisms.

B. Datasets

The experimental evaluation was designed to isolate the relationships among graph topology, diffusion dynamics, and degree-biased transport behavior. For that, we selected datasets that exhibit fundamentally different structural characteristics.

The *Cora* citation network [24] serves as the primary homophilic benchmark, where neighboring nodes frequently share similar class labels. To evaluate sparse heterophily, we utilized the *Texas*, *Wisconsin*, and *Cornell* WebKB webpage networks [11,25]. These graphs contain highly sparse hyperlink structures with strong cross-class connectivity. For dense heterophilic regimes, we evaluated the model on the *Chameleon* and *Squirrel* Wikipedia article networks [26], both characterized by dense connectivity, heavy-tailed degree distributions, and dominant structural hubs typical of scale-free organization. Finally, the *Actor* co-occurrence network [27] was included as a non-scale-free control graph to evaluate whether the proposed routing mechanism remains specifically coupled to scale-free transport dynamics. The WebKB, Wikipedia, and Actor benchmark variants used in this work correspond to the formulations introduced by Pei *et al.* [11] for evaluating graph neural networks under heterophily.

Further details of the datasets are summarized in Table I, where the homophily ratio denotes the fraction of edges connecting nodes sharing identical class labels. Values approaching 1 indicate a strong homophilic structure, whereas values near 0 indicate highly heterophilic connectivity.

C. Graph preprocessing pipeline

All datasets were implemented using the PyTorch Geometric framework [28]. The edge-index representation of each graph was converted into a sparse adjacency matrix before the construction of the PDRW transition operator.

TABLE I. Summary of benchmark datasets and their structural characteristics.

Dataset	Network type	Nodes	Edges	Features	Classes	Homophily ratio
Cora	Citation network	2708	5429	1433	7	0.81
Texas	WebKB webpages	183	309	1703	5	0.11
Wisconsin	WebKB webpages	251	499	1703	5	0.21
Cornell	WebKB webpages	183	298	1703	5	0.30
Chameleon	Wikipedia pages	2277	31 421	2325	5	0.23
Squirrel	Wikipedia pages	5201	198 493	2089	5	0.22
Actor	Actor co-occurrence	7600	26 659	931	5	0.22

As stated before, to avoid numerical instability under negative values of β , isolated nodes with degree $k_i = 0$ are assigned zero transition weight, preventing undefined degree exponentiation.

For all experiments, self-loops were introduced through the adaptive stabilization strategy. The resulting routing matrix was subsequently row normalized to produce the final stochastic transition operator. This normalization ensures probabilistic conservation during message propagation and stabilizes diffusion dynamics across multiple graph convolutional layers.

D. Training configuration

To isolate the effect of topology-dependent diffusion from architectural variation, all experiments utilized a fixed neural network configuration across datasets.

The PDRW-GCN architecture consisted of two graph convolutional layers with a hidden dimension of 32 and ReLU activation. Model optimization was performed using the Adam optimizer with learning rate $\alpha = 0.01$ and weight decay 5×10^{-4} . A dropout rate of 0.5 was applied after the hidden layer to reduce overfitting, particularly on the small WebKB graphs.

All models were trained for 200 epochs using cross-entropy loss. Dataset partitions followed fixed semisupervised splits consisting of 48% training nodes, 32% validation nodes, and 20% testing nodes. The fixed hyperparameter configuration shown in Table II was maintained across all datasets to isolate the effect of topology-dependent diffusion from architectural variation.

To reduce stochastic fluctuations arising from random initialization and sparse graph sampling, all reported performance metrics were averaged over ten independent training trials. This averaging procedure was particularly important for the small WebKB datasets, where individual training runs often exhibit large variance in validation accuracy due to limited neighborhood connectivity and strong heterophily.

TABLE II. Uniform hyperparameter configuration used across all experiments.

Hyperparameter	Value
Optimizer	Adam
Learning rate	0.01
Weight decay	5×10^{-4}
Hidden dimension	32
Dropout rate	0.5
Training epochs	200
Number of layers	2

E. Evaluation metrics

Model performance was evaluated using accuracy, precision, recall, F1 score, and ROC AUC. Because the sparse heterophilic datasets exhibit substantial class imbalance, relying solely on classification accuracy can produce misleading estimates of model quality. The inclusion of F1 score and ROC AUC therefore provides a more robust assessment of latent class separability under different topology-dependent diffusion regimes.

Accuracy measures the fraction of correctly classified nodes, while precision and recall quantify the false-positive and false-negative rates, respectively. The F1 score captures the harmonic balance between precision and recall, and ROC AUC evaluates the separability of the learned latent representations across classification thresholds.

F. Computational environment

All experiments were implemented in Python using PyTorch [29] and PyTorch geometric [28]. Preliminary graph analysis and visualization were performed using NetworkX, while latent-space visualizations used t-SNE projections [30]. Training and evaluation were accelerated using CUDA-enabled graphics processing unit (GPU) within the Google Colab environment.

G. Bias parameter selection

The structural bias parameter β was evaluated over the interval $[-2.0, 2.0]$. This range was deliberately selected to capture both strong hub amplification ($\beta > 0$) and suppression ($\beta < 0$) while operating strictly within the nondegenerate regime of degree-biased random walks. Mechanistically, internode transition probabilities scale as $P_{ij} \propto k_j^\beta$. As $|\beta|$ increases beyond 2.0, these probabilities begin to saturate, gradually converting the stochastic diffusion process into a deterministic routing operator. For $\beta > 2.0$, transition weights become overwhelmingly dominated by the single highest-degree neighbor within a local neighborhood, effectively ignoring alternative structural pathways. Conversely, for $\beta < -2.0$, message passing collapses into deterministic routing toward low-degree leaf nodes ($k_j = 1$). The interval $[-2.0, 2.0]$ thus defines the operational bounds where degree-driven routing preferences are strongly asserted without collapsing multipath diffusion dynamics or introducing numerical instability.

In the present study, β was intentionally evaluated as a fixed control parameter across this interval to isolate the direct mechanistic influence of topology-dependent diffusion bias on graph propagation dynamics. As reported in Table III, extending β beyond this operational interval reveals a pronounced topology-dependent divergence in routing behavior. In dense Wikipedia-based networks (Chameleon and Squirrel), strongly negative bias beyond $\beta = -2.0$ rapidly degrades performance by excessively suppressing structural hubs, effectively fragmenting the graph’s primary transport pathways and isolating local neighborhoods.

In contrast, sparse WebKB datasets exhibit an unusual apparent resilience under large positive values of β . These networks contain substantial structural noise and severe class imbalance, such

TABLE III. Effect of extreme bias (β) magnitudes on validation accuracy.

Dataset	$ \beta = 2.0(\%)$	$ \beta = 5.0(\%)$	$ \beta = 10.0(\%)$	$ \beta = 20.0(\%)$	$ \beta = 40.0(\%)$	$ \beta = 80.0(\%)$
Wikipedia networks (evaluated at $\beta < 0$)						
Chameleon	66.78 ± 1.71	66.86 ± 2.27	66.21 ± 1.63	64.89 ± 2.37	48.36 ± 2.27	50.09 ± 2.15
Squirrel	53.80 ± 2.04	54.06 ± 2.00	53.15 ± 2.72	51.73 ± 2.13	35.12 ± 1.00	34.06 ± 1.50
WebKB networks (evaluated at $\beta > 0$)						
Texas	78.11 ± 4.90	79.19 ± 5.93	77.03 ± 4.23	15.68 ± 4.80	15.68 ± 4.80	15.68 ± 4.80
Cornell	67.84 ± 5.33	70.54 ± 4.90	66.76 ± 8.72	24.05 ± 6.22	24.05 ± 6.22	24.05 ± 6.22
Wisconsin	75.69 ± 5.13	78.04 ± 4.28	79.22 ± 5.20	3.73 ± 1.85	3.73 ± 1.85	3.73 ± 1.85

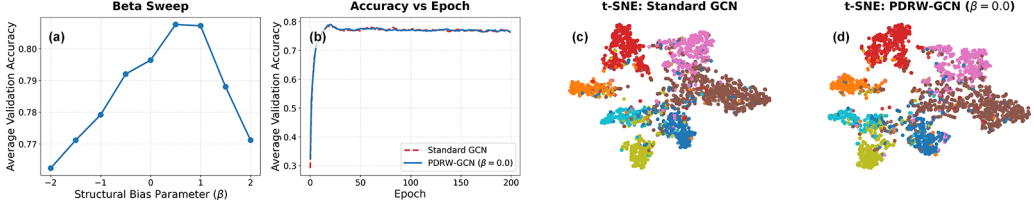


FIG. 1. Performance analysis of PDRW-GCN on the homophilic Cora dataset. From left to right: (a) validation accuracy as a function of the structural bias parameter β , (b) validation accuracy during training for standard GCN and PDRW-GCN, and (c), (d) t-SNE visualizations of the learned latent representations for the standard GCN and PDRW-GCN, respectively.

that concentrating diffusion toward a small subset of dominant hubs partially decouples neighborhood aggregation from the surrounding noisy topology. The propagation dynamics increasingly resemble localized feature transformations akin to a multilayer perceptron, operating with diminished reliance on broader topological connectivity.

This regime, however, represents an artifact of topological degeneration rather than effective graph learning. Once β reaches extreme magnitudes ($|\beta| \gg 2.0$), the transition operator numerically saturates, causing diffusion to collapse onto an extremely restricted subset of nodes. Consequently, the routing dynamics lose expressive variability, and model performance asymptotically converges toward a near-constant baseline. These empirical observations indicate that $\beta \in [-2.0, 2.0]$ captures the true nondegenerate operational spectrum of degree-biased message passing.

IV. RESULTS AND DISCUSSION

In this section, we investigate how topology-dependent diffusion bias influences graph message passing across different structural regimes. The analysis is organized according to graph topology, heterophily characteristics, and the resulting diffusion-transport behavior.

A. Homophilic regime

We first evaluate the proposed architecture on the homophilic Cora citation network, where neighboring nodes frequently share the same semantic class. In such systems, standard neighborhood averaging already produces highly coherent latent representations due to strong local diffusion consistency.

Consistent with this expectation, strong performance on Cora is concentrated in a broad high-accuracy regime near unbiased diffusion dynamics. As shown in the structural bias sweep in Fig. 1(a), average validation accuracy exhibits a smooth, concave profile that forms a plateau between $\beta = 0.0$ and $\beta = 1.0$ (peaking at $\approx 80.7\%$ around $\beta = 0.5$ – 1.0), before falling off toward both extreme negative ($\beta = -2.0$, $\approx 76.3\%$) and extreme positive ($\beta = +2.0$, $\approx 77.1\%$) bias regimes. In a homophilic graph, where local neighbors share consistent semantic labels, mild degree weighting ($\beta \in [0, 1]$) preserves stable, high-fidelity feature aggregation. However, excessive positive bias ($\beta \geq 2.0$) concentrates aggregation too aggressively onto high-degree citation hubs, inducing local feature oversmoothing. Conversely, strong negative bias ($\beta \leq -2.0$) suppresses structurally critical hubs, artificially fragmenting neighborhood connectivity and weakening long-range semantic propagation.

The additional evaluation metrics reported in Table IV further confirm that near-unbiased diffusion remains optimal in homophilic environments. PDRW-GCN at $\beta = 0.0$ achieves a test accuracy of 87.56%, an F1 score of 87.58%, and an ROC AUC of 98.20%, outperforming standard baseline architectures such as standard GCN (87.04% accuracy, 87.05% F1) and GAT (86.43% accuracy, 86.40% F1). The training dynamics in Fig. 1(b) demonstrate that both Standard GCN and PDRW-GCN ($\beta = 0.0$) rapidly converge within 20 epochs and maintain stable, overlapping

TABLE IV. Node classification performance (mean $\pm$ std % across ten random/standard splits) across homophilic, heterophilic, and non-scale-free graph regimes. Best results are shown in bold.

Dataset	Model	Test accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	ROC AUC (%)
Cora	Standard GCN	87.04 $\pm$ 1.92	87.36 $\pm$ 1.93	87.04 $\pm$ 1.92	87.05 $\pm$ 1.90	97.95 $\pm$ 0.56
	GAT	86.43 $\pm$ 2.40	86.75 $\pm$ 2.18	86.43 $\pm$ 2.40	86.40 $\pm$ 2.39	97.97 $\pm$ 0.56
	CenGCN	85.31 $\pm$ 2.57	85.60 $\pm$ 2.39	85.31 $\pm$ 2.57	85.27 $\pm$ 2.57	97.73 $\pm$ 0.69
	FAGCN	86.91 $\pm$ 3.46	87.43 $\pm$ 2.93	86.91 $\pm$ 3.46	86.92 $\pm$ 3.38	98.01 $\pm$ 0.72
	GPR-GNN	86.74 $\pm$ 2.84	87.26 $\pm$ 2.14	86.74 $\pm$ 2.84	86.81 $\pm$ 2.72	98.11 $\pm$ 0.59
	Geom-GCN	86.48 $\pm$ 2.96	86.85 $\pm$ 2.61	86.48 $\pm$ 2.96	86.48 $\pm$ 2.94	97.55 $\pm$ 0.80
	H2GCN	87.15 $\pm$ 2.20	87.51 $\pm$ 1.91	87.15 $\pm$ 2.20	87.18 $\pm$ 2.13	98.16 $\pm$ 0.57
	PDRW-GCN ($\beta = 0.0$)	87.56 $\pm$ 1.93	87.84 $\pm$ 1.79	87.56 $\pm$ 1.93	87.58 $\pm$ 1.91	98.20 $\pm$ 0.65
Texas	Standard GCN	56.76 $\pm$ 4.01	40.36 $\pm$ 7.95	56.76 $\pm$ 4.01	46.32 $\pm$ 5.66	48.58 $\pm$ 3.74
	GAT	57.57 $\pm$ 5.41	44.07 $\pm$ 7.18	57.57 $\pm$ 5.41	48.56 $\pm$ 5.33	51.08 $\pm$ 4.26
	CenGCN	67.84 $\pm$ 7.49	62.41 $\pm$ 11.37	67.84 $\pm$ 7.49	63.59 $\pm$ 9.48	54.82 $\pm$ 9.67
	FAGCN	60.81 $\pm$ 5.16	47.50 $\pm$ 8.16	60.81 $\pm$ 5.16	51.49 $\pm$ 6.43	51.06 $\pm$ 7.50
	GPR-GNN	71.35 $\pm$ 9.83	71.18 $\pm$ 9.18	71.35 $\pm$ 9.83	69.82 $\pm$ 9.68	58.49 $\pm$ 16.99
	Geom-GCN	59.46 $\pm$ 4.98	45.71 $\pm$ 11.07	59.46 $\pm$ 4.98	48.86 $\pm$ 7.77	52.11 $\pm$ 4.72
	H2GCN	67.84 $\pm$ 8.84	63.49 $\pm$ 12.05	67.84 $\pm$ 8.84	63.24 $\pm$ 11.22	56.76 $\pm$ 13.73
	PDRW-GCN ($\beta = 2.0$)	78.11 $\pm$ 6.78	73.87 $\pm$ 10.00	78.11 $\pm$ 6.78	74.84 $\pm$ 8.59	58.16 $\pm$ 16.32
Wisconsin	Standard GCN	51.57 $\pm$ 4.48	49.93 $\pm$ 9.02	51.57 $\pm$ 4.48	48.63 $\pm$ 5.96	66.45 $\pm$ 4.89
	GAT	52.94 $\pm$ 5.33	44.40 $\pm$ 4.47	52.94 $\pm$ 5.33	46.39 $\pm$ 5.63	64.28 $\pm$ 3.68
	CenGCN	63.73 $\pm$ 5.96	60.61 $\pm$ 7.79	63.73 $\pm$ 5.96	60.62 $\pm$ 6.61	80.35 $\pm$ 3.84
	FAGCN	65.88 $\pm$ 4.40	57.38 $\pm$ 4.37	65.88 $\pm$ 4.40	58.95 $\pm$ 5.31	80.69 $\pm$ 3.93
	GPR-GNN	79.22 $\pm$ 5.76	78.95 $\pm$ 4.88	79.22 $\pm$ 5.76	78.05 $\pm$ 5.44	93.02 $\pm$ 3.12
	Geom-GCN	53.53 $\pm$ 4.64	46.96 $\pm$ 4.93	53.53 $\pm$ 4.64	47.76 $\pm$ 5.57	63.97 $\pm$ 5.05
	H2GCN	75.69 $\pm$ 5.63	76.17 $\pm$ 5.48	75.69 $\pm$ 5.63	74.63 $\pm$ 5.10	90.55 $\pm$ 4.69
	PDRW-GCN ($\beta = 2.0$)	76.08 $\pm$ 3.59	74.87 $\pm$ 5.98	76.08 $\pm$ 3.59	73.02 $\pm$ 4.49	91.76 $\pm$ 3.49
Cornell	Standard GCN	45.14 $\pm$ 6.95	45.76 $\pm$ 9.49	45.14 $\pm$ 6.95	43.38 $\pm$ 6.83	66.59 $\pm$ 6.28
	GAT	44.32 $\pm$ 6.64	37.13 $\pm$ 11.36	44.32 $\pm$ 6.64	37.70 $\pm$ 8.89	61.67 $\pm$ 7.01
	CenGCN	58.11 $\pm$ 5.44	58.47 $\pm$ 5.89	58.11 $\pm$ 5.44	53.86 $\pm$ 5.57	78.31 $\pm$ 3.10
	FAGCN	50.81 $\pm$ 5.10	47.91 $\pm$ 11.31	50.81 $\pm$ 5.10	43.80 $\pm$ 6.66	77.93 $\pm$ 4.77
	GPR-GNN	65.95 $\pm$ 5.57	66.30 $\pm$ 5.27	65.95 $\pm$ 5.57	64.95 $\pm$ 4.64	85.38 $\pm$ 4.01
	Geom-GCN	45.14 $\pm$ 5.41	40.34 $\pm$ 10.96	45.14 $\pm$ 5.41	40.00 $\pm$ 7.62	62.62 $\pm$ 5.81
	H2GCN	61.35 $\pm$ 6.40	61.24 $\pm$ 5.64	61.35 $\pm$ 6.40	58.83 $\pm$ 6.46	83.51 $\pm$ 4.84
	PDRW-GCN ($\beta = 2.0$)	67.57 $\pm$ 4.68	66.72 $\pm$ 8.29	67.57 $\pm$ 4.68	64.69 $\pm$ 6.70	86.75 $\pm$ 3.80
Chameleon	Standard GCN	64.14 $\pm$ 2.27	66.08 $\pm$ 2.57	64.14 $\pm$ 2.27	64.41 $\pm$ 2.27	86.19 $\pm$ 1.01
	GAT	43.62 $\pm$ 2.70	43.65 $\pm$ 3.02	43.62 $\pm$ 2.70	42.51 $\pm$ 2.57	74.42 $\pm$ 1.89
	CenGCN	66.69 $\pm$ 1.64	67.47 $\pm$ 1.78	66.69 $\pm$ 1.64	66.76 $\pm$ 1.64	88.07 $\pm$ 0.63
	FAGCN	56.01 $\pm$ 2.21	56.86 $\pm$ 2.25	56.01 $\pm$ 2.21	55.95 $\pm$ 2.27	83.63 $\pm$ 0.88
	GPR-GNN	61.32 $\pm$ 1.63	61.94 $\pm$ 1.83	61.32 $\pm$ 1.63	61.26 $\pm$ 1.67	85.58 $\pm$ 0.92
	Geom-GCN	65.33 $\pm$ 1.56	66.85 $\pm$ 1.93	65.33 $\pm$ 1.56	65.44 $\pm$ 1.53	86.09 $\pm$ 1.05
	H2GCN	57.43 $\pm$ 2.24	57.96 $\pm$ 2.39	57.43 $\pm$ 2.24	57.43 $\pm$ 2.23	83.45 $\pm$ 0.92
	PDRW-GCN ($\beta = -2.0$)	66.78 $\pm$ 2.01	67.35 $\pm$ 1.91	66.78 $\pm$ 2.01	66.84 $\pm$ 1.99	87.08 $\pm$ 0.94
Squirrel	Standard GCN	47.73 $\pm$ 1.38	47.99 $\pm$ 1.66	47.73 $\pm$ 1.38	46.60 $\pm$ 1.67	76.52 $\pm$ 0.75
	GAT	28.77 $\pm$ 1.28	28.83 $\pm$ 1.22	28.77 $\pm$ 1.28	28.29 $\pm$ 1.40	59.22 $\pm$ 0.90
	CenGCN	51.71 $\pm$ 1.97	51.81 $\pm$ 1.86	51.71 $\pm$ 1.97	51.27 $\pm$ 2.02	77.12 $\pm$ 0.89
	FAGCN	39.92 $\pm$ 1.18	39.74 $\pm$ 1.22	39.92 $\pm$ 1.18	39.66 $\pm$ 1.19	71.43 $\pm$ 0.71
	GPR-GNN	42.67 $\pm$ 1.53	43.71 $\pm$ 1.40	42.67 $\pm$ 1.53	42.34 $\pm$ 1.42	73.52 $\pm$ 1.33
	Geom-GCN	48.57 $\pm$ 1.61	48.98 $\pm$ 1.22	48.57 $\pm$ 1.61	48.10 $\pm$ 1.51	75.50 $\pm$ 0.92
	H2GCN	37.25 $\pm$ 1.62	37.02 $\pm$ 1.76	37.25 $\pm$ 1.62	36.91 $\pm$ 1.63	67.87 $\pm$ 0.98
	PDRW-GCN ($\beta = -2.0$)	54.00 $\pm$ 1.88	53.89 $\pm$ 1.88	54.00 $\pm$ 1.88	53.79 $\pm$ 1.87	76.64 $\pm$ 1.05

TABLE IV. (Continued.)

Dataset	Model	Test accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	ROC AUC (%)
Actor	Standard GCN	30.22 ± 0.61	23.08 ± 2.46	30.22 ± 0.61	25.00 ± 0.95	59.22 ± 0.30
	GAT	28.17 ± 1.46	26.65 ± 4.07	28.17 ± 1.46	22.12 ± 1.66	57.77 ± 0.50
	CenGCN	29.88 ± 1.28	25.10 ± 3.68	29.88 ± 1.28	25.06 ± 1.96	60.41 ± 0.66
	FAGCN	33.41 ± 0.74	29.80 ± 1.36	33.41 ± 0.74	31.03 ± 1.24	65.43 ± 0.77
	GPR-GNN	35.23 ± 1.01	35.99 ± 2.02	35.23 ± 1.01	34.29 ± 1.57	68.37 ± 0.41
	Geom-GCN	29.80 ± 0.96	22.37 ± 1.98	29.80 ± 0.96	24.63 ± 1.05	58.22 ± 0.65
	H2GCN	34.88 ± 0.95	33.02 ± 3.06	34.88 ± 0.95	32.76 ± 1.43	68.20 ± 0.67
	PDRW-GCN ($\beta = 0.0$)	29.36 ± 1.16	23.54 ± 2.38	29.36 ± 1.16	23.81 ± 1.91	59.06 ± 0.75

validation trajectories ($\approx 77.5 - 78.0\%$). Furthermore, comparing the t-SNE visualizations reveals that standard GCN [Fig. 1(c)] and PDRW-GCN [$\beta = 0.0$, Fig. 1(d)] produce virtually identical well-separated class clusters, indicating that homophilic networks do not require degree-biased routing adjustments.

B. Sparse heterophilic regime

Next, we examine sparse heterophilic graphs using the WebKB datasets: Texas [Fig. 2(a)], Wisconsin [Fig. 2(b)], and Cornell [Fig. 2(c)]. These networks feature sparse hyperlink

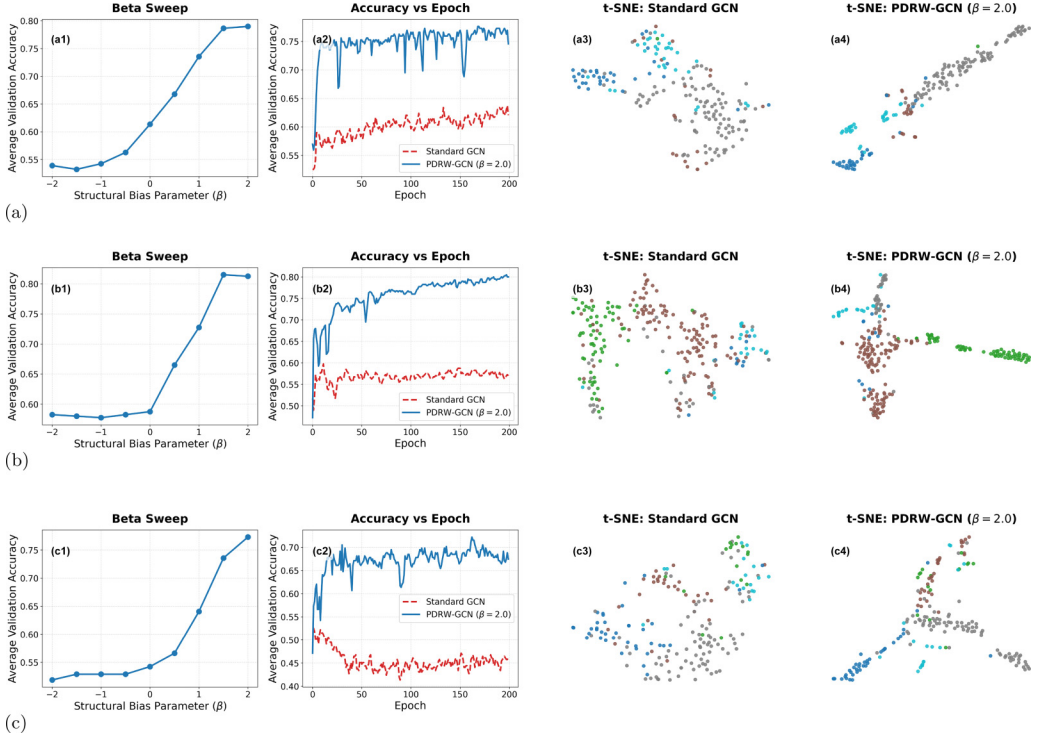


FIG. 2. Performance analysis of the sparse heterophilic WebKB datasets: (a) Texas, (b) Wisconsin, and (c) Cornell. For each dataset, the panels show (a1)–(c1) validation accuracy across the structural bias range, (a2)–(c2) validation accuracy during training for standard GCN and PDRW-GCN, and (a3)–(c3), (a4)–(c4) t-SNE visualizations of the corresponding latent embeddings.

topologies dominated by cross-class connectivity, causing standard neighborhood averaging to conflate incompatible semantic signals.

Under these conditions, positive structural bias ($\beta > 0$) substantially improves node classification performance by preferentially routing information through high-degree transport hubs. In the Texas bias sweep [Fig. 2(a1)], validation accuracy remains low under negative bias ($\beta = -2.0$, $\approx 28.5\%$) and unbiased routing ($\beta = 0.0$, $\approx 42.5\%$), but climbs as β increases, peaking near $\beta = 1.5$ ($\approx 80.0\%$) and $\beta = 2.0$ ($\approx 78.7\%$). Similar steep performance gains occur on Wisconsin [Fig. 2(b1)], rising from $\approx 48.5\%$ at $\beta = 0.0$ to a peak of $\approx 72.5\%$ near $\beta = 1.5$ – 2.0 , and on Cornell [Fig. 2(c1)], where validation accuracy jumps from $\approx 34.0\%$ at $\beta = 0.0$ to over 63.5% at $\beta = 2.0$.

The epoch convergence plots further highlight the failure of isotropic message passing in sparse heterophilic topologies. On Texas [Fig. 2(a2)], standard GCN stagnates between 55% and 64% validation accuracy, whereas PDRW-GCN ($\beta = 2.0$) rapidly climbs past 75% by epoch 10 and maintains a consistently high trajectory ($\approx 70\%$ – 78%). On Wisconsin [Fig. 2(b2)], standard GCN remains flat near 55% – 58% , while PDRW-GCN ($\beta = 2.0$) steadily ascends from epoch 0 to reach nearly 80% validation accuracy. On Cornell [Fig. 2(c2)], standard GCN degrades from 50% down to 40% – 45% , while PDRW-GCN ($\beta = 2.0$) rapidly settles at 65% – 71% .

This performance recovery is strongly reflected across all evaluation metrics in Table IV. On Texas, PDRW-GCN ($\beta = 2.0$) leads all models with 78.11% accuracy and 74.84% F1 score (compared to standard GCN’s 56.76% accuracy and 46.32% F1 score). On Cornell, PDRW-GCN ($\beta = 2.0$) achieves overall top performance with 67.57% accuracy and 86.75% ROC AUC. On Wisconsin, PDRW-GCN ($\beta = 2.0$) achieves 76.08% accuracy and 91.76% ROC AUC, closely trailing GPR-GNN (79.22%).

The qualitative impact of positive hub routing is strikingly evident in the t-SNE projections. Standard GCN on Texas [Fig. 2(a3)], Wisconsin [Fig. 2(b3)], and Cornell [Fig. 2(c3)] yields disorganized, overlapping point clouds. In contrast, PDRW-GCN with positive bias [$\beta = 2.0$, Figs. 2(a4), 2(b4), and 2(c4)] reorganizes the feature space into comparatively distinct class clusters and linear branching manifold structures.

C. Dense heterophilic regime

The model’s behavior shifts significantly in dense heterophilic graphs, such as Chameleon [Fig. 3(a)] and Squirrel [Fig. 3(b)]. Unlike sparse WebKB datasets, these networks exhibit high local node connectivity alongside prominent hub structures, making standard feature aggregation particularly susceptible to oversmoothing and cross-class feature mixing.

Under positive structural bias ($\beta > 0$), routing is directed heavily toward high-degree nodes, which further amplifies feature mixing across semantically inconsistent neighborhoods. Consequently, performance degrades steeply as β increases. On Chameleon [Fig. 3(a1)], validation accuracy falls from $\approx 66.5\%$ at $\beta = -2.0$ down to $\approx 43.5\%$ at $\beta = +2.0$. Similarly, on Squirrel [Fig. 3(b1)], performance plummets from $\approx 54.0\%$ at $\beta = -2.0$ to $\approx 32.3\%$ at $\beta = +2.0$. Suppressing hub transitions ($\beta < 0$) mitigates this noise broadcast, yielding optimal performance under negative bias.

However, compared to standard GCN, the performance advantage provided by negative bias remains relatively modest, highlighting the intrinsic difficulty of dense heterophilic topologies. On Chameleon [Fig. 3(a2)], PDRW-GCN ($\beta = -2.0$) offers only an incremental improvement over standard GCN ($\approx 67.5\%$ versus $\approx 65.0\%$ validation accuracy). On Squirrel [Fig. 3(b2)], the margin is somewhat more pronounced ($\approx 54.5\%$ versus $\approx 47.0\%$). While Table IV reflects competitive overall results (66.78% accuracy on Chameleon and 54.00% accuracy on Squirrel), negative structural bias functions primarily as a local noise dampener rather than a complete cure for dense neighborhood mixing.

This structural limitation is visually evident in the t-SNE projections. Unlike the sparse heterophilic regime where clear cluster boundaries formed, the latent embeddings for both standard GCN and PDRW-GCN on Chameleon [Figs. 3(a3) and 3(a4)] and Squirrel [Figs. 3(b3) and 3(b4)]

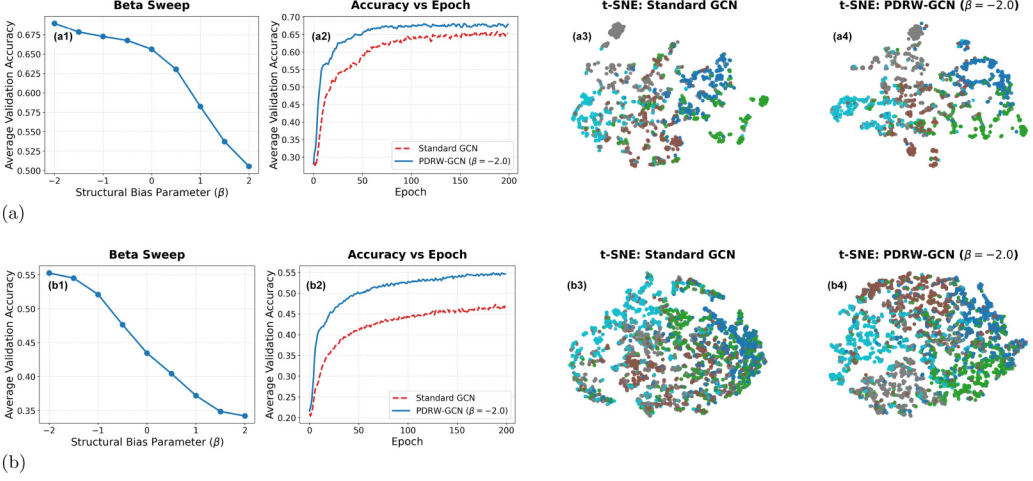


FIG. 3. Performance analysis of the dense heterophilic Wikipedia networks: (a) Chameleon and (b) Squirrel. For each dataset, the panels show [(a1), (b1)] validation accuracy as a function of β , [(a2), (b2)] validation accuracy during training for standard GCN and PDRW-GCN, and [(a3), (b3)], [(a4), (b4)] t-SNE projections of the learned latent representations.

retain substantial class overlap and intermingling. While hub suppression ($\beta = -2.0$) subtly reduces background point dispersion and preserves localized class density, the global feature space remains largely entangled. This indicates that while negative degree bias successfully contains hub-driven noise propagation, dense heterophily presents fundamental structural boundaries that local topological routing alone cannot completely untangle.

D. Non-scale-free control graph

To verify whether the proposed mechanism acts as a genuine topology-dependent routing process rather than a generic regularizer, we evaluated PDRW-GCN on the Actor co-occurrence network (Fig. 4), which lacks a strongly scale-free degree distribution.

As shown in the Actor bias sweep [Fig. 4(a)], average validation accuracy remains constrained within a narrow band across all parameter values (between 25.7% and 29.7%). Although a minor bump occurs near $\beta = +0.5$ ($\approx 29.7\%$), performance degrades under strong positive bias ($\beta = +2.0$, 25.7%) and negative bias ($\beta = -2.0$, 27.0%). In Fig. 4(b), both Standard GCN and PDRW-GCN ($\beta = 0.0$) follow almost identical training trajectories, peaking briefly around epoch 30–40 ($\approx 31.5\%$) before slowly decaying to $\approx 29.5\%$ due to feature mixing and oversmoothing.

The benchmark results in Table IV further highlight the limitations of local degree-biased routing on non-scale-free graphs: Local message-passing baselines (standard GCN 30.22%,

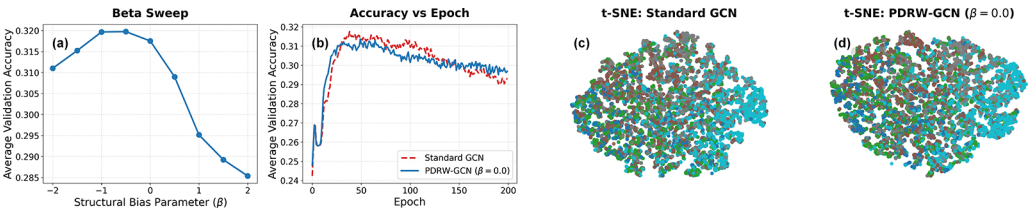


FIG. 4. Performance analysis of the non-scale-free Actor co-occurrence network. From left to right: (a) validation accuracy across the structural bias range, (b) validation accuracy during training for standard GCN and PDRW-GCN, and (c), (d) t-SNE visualizations of the resulting latent embeddings.

PDRW-GCN 29.36%) perform noticeably worse than higher-order adaptive architectures like GPR-GNN (35.23%) and H2GCN (34.88%).

The absence of effective topological routing on Actor is most evident in the t-SNE projections. Both standard GCN [Fig. 4(c)] and PDRW-GCN [$\beta = 0.0$, Fig. 4(d)] produce dense, completely overlapping spherical point clouds where classes remain indistinguishable. This suggests that PDRW-GCN’s degree-biased routing mechanism requires a scale-free heavy-tailed degree hierarchy to operate effectively; in non-scale-free topologies lacking clear hub structures, biased random walks offer no structural advantage.

These results further suggest that the effectiveness of the proposed routing operator depends critically on heavy-tailed degree organization and the existence of structurally dominant hubs governing large-scale transport pathways.

E. Comparative evaluation across graph regimes

Table IV reports the node classification results across seven datasets representing homophilic, heterophilic, and dense graph structures. On the homophilic Cora dataset, PDRW-GCN ($\beta = 0.0$) achieves the highest performance across all evaluation metrics, reaching a test accuracy of 87.56% and an F1 score of 87.58%, slightly outperforming both standard GCN and specialized heterophily models like H2GCN.

On sparse heterophilic web graphs (Texas, Wisconsin, and Cornell), PDRW-GCN ($\beta = +2.0$) secures the best test accuracy on Texas (78.11%) and Cornell (67.57%), showing up to a 21.35% accuracy improvement over standard GCN. On Wisconsin, while GPR-GNN obtains the top performance (79.22%), PDRW-GCN remains highly competitive at 76.08%.

For dense heterophilic networks (Chameleon and Squirrel), PDRW-GCN ($\beta = -2.0$) achieves top classification accuracy on both Chameleon (66.78%) and Squirrel (54.00%), outperforming dedicated heterophily architectures such as FAGCN and Geom-GCN. On the Actor dataset, GPR-GNN leads overall with 35.23% accuracy, whereas PDRW-GCN performs comparably to standard GCN at 29.36%. Overall, these empirical results confirm that tuning the structural bias parameter β allows the network to adapt effectively across varying edge densities and homophily levels without adding extra parameter complexity.

F. Ablation analysis

To isolate the mechanistic contributions of the proposed stabilization and routing components, we performed a series of ablation experiments examining the joint effects of self-loop scaling, structural bias parameter range, and propagation depth. These studies determine whether the observed accuracy gains stem from genuine topology-dependent diffusion behavior rather than incidental architectural regularization. The corresponding quantitative results are summarized in Tables V and VI.

1. Self-loop multiplier

To quantify the balance between ego-feature preservation and neighborhood aggregation, we evaluated the joint effect of varying the self-loop amplification factor w_{self} and structural bias parameter β across representative homophilic and heterophilic datasets.

It is necessary to distinguish the mechanistic role of the self-loop multiplier w_{self} from the structural bias parameter β . While increasing w_{self} scales the diagonal of the adjacency matrix to enhance ego-feature retention and dampen cross-class feature mixing, it does not alter the relative weighting among external neighbors. In contrast, the structural bias parameter β fundamentally reweights internode transition probabilities P_{ij} , controlling which neighboring structural profiles (hubs or low-degree nodes) dominate message passing. Thus, ego-feature retention (w_{self}) and degree-biased routing (β) operate as “orthogonal” mechanisms: w_{self} controls the balance between self-information and total neighborhood context, whereas β determines the directional selectivity within the neighborhood context itself.

TABLE V. Joint ablation study across structural bias (β) and self-loop amplification (w_{self}). Validation accuracy (mean $\pm$ std dev) across ten random splits on Cora (homophilic), Wisconsin (sparse heterophilic), and Chameleon (dense heterophilic). Best values are shown in bold.

Structural Bias (β)	Self-loop multiplier (w_{self})					
	1.0 $\times$ (%)	2.0 $\times$ (%)	4.0 $\times$ (%)	6.0 $\times$ (%)	8.0 $\times$ (%)	10.0 $\times$ (%)
Cora (homophilic benchmark)						
-2.0	83.12 $\pm$ 2.69	83.25 $\pm$ 2.71	83.38 $\pm$ 3.59	82.70 $\pm$ 3.62	82.75 $\pm$ 3.60	82.43 $\pm$ 3.50
-1.0	85.73 $\pm$ 2.64	85.22 $\pm$ 2.90	85.26 $\pm$ 3.06	84.80 $\pm$ 3.36	84.64 $\pm$ 3.46	84.36 $\pm$ 3.16
0.0	87.36 $\pm$ 1.50	86.87 $\pm$ 2.57	87.07 $\pm$ 2.78	86.04 $\pm$ 3.24	85.22 $\pm$ 3.16	85.80 $\pm$ 3.43
+1.0	85.70 $\pm$ 2.33	84.45 $\pm$ 2.26	84.64 $\pm$ 2.69	83.05 $\pm$ 2.09	83.35 $\pm$ 3.16	82.25 $\pm$ 2.60
+2.0	80.46 $\pm$ 3.03	79.83 $\pm$ 3.89	78.79 $\pm$ 4.37	77.46 $\pm$ 4.01	76.33 $\pm$ 3.21	76.33 $\pm$ 4.30
Wisconsin (sparse heterophilic benchmark)						
-2.0	50.20 $\pm$ 4.04	51.57 $\pm$ 4.39	54.71 $\pm$ 4.76	59.22 $\pm$ 4.19	61.76 $\pm$ 3.42	65.88 $\pm$ 3.53
-1.0	51.76 $\pm$ 4.90	48.04 $\pm$ 4.66	52.75 $\pm$ 2.83	57.65 $\pm$ 3.30	60.20 $\pm$ 3.17	60.59 $\pm$ 3.22
0.0	49.02 $\pm$ 4.11	51.57 $\pm$ 3.62%	54.31 $\pm$ 3.04	58.82 $\pm$ 2.91	60.78 $\pm$ 3.82	63.73 $\pm$ 3.85
+1.0	62.75 $\pm$ 3.40	63.92 $\pm$ 6.15	72.55 $\pm$ 4.88	79.22 $\pm$ 5.20	79.61 $\pm$ 5.63	82.16 $\pm$ 4.84
+2.0	71.57 $\pm$ 4.23	76.47 $\pm$ 4.21	83.14 $\pm$ 4.13	83.53 $\pm$ 4.57	84.31 $\pm$ 3.04	84.51 $\pm$ 3.66
Chameleon (dense heterophilic benchmark)						
-2.0	68.99 $\pm$ 2.29	67.11 $\pm$ 1.92	64.80 $\pm$ 1.77	63.05 $\pm$ 1.43	60.96 $\pm$ 1.43	60.29 $\pm$ 1.61
-1.0	67.65 $\pm$ 1.46	65.35 $\pm$ 2.19	62.24 $\pm$ 1.81	60.64 $\pm$ 1.80	59.28 $\pm$ 1.36	57.61 $\pm$ 1.03
0.0	63.42 $\pm$ 1.87	62.50 $\pm$ 2.44	60.13 $\pm$ 2.76	57.68 $\pm$ 1.82	55.46 $\pm$ 1.83	54.80 $\pm$ 2.25
+1.0	52.68 $\pm$ 2.49	53.86 $\pm$ 2.24	55.09 $\pm$ 1.53	55.86 $\pm$ 2.24	55.46 $\pm$ 2.33	54.65 $\pm$ 1.57
+2.0	43.00 $\pm$ 3.23	43.68 $\pm$ 2.30	48.03 $\pm$ 2.30	48.55 $\pm$ 3.38	49.19 $\pm$ 2.40	49.63 $\pm$ 1.92

In our joint sensitivity experiments across $w_{\text{self}} \in \{1.0, 2.0, 4.0, 6.0, 8.0, 10.0\}$ and $\beta \in \{-2.0, -1.0, 0.0, 1.0, 2.0\}$, we observe two key empirical insights (summarized in Table V). First is that the optimal sign of β remains invariant across all values of w_{self} . On sparse heterophilic graphs (Wisconsin), positive structural bias ($\beta = +2.0$) consistently yields superior performance for every fixed self-loop scale. Conversely, on dense heterophilic graphs (Chameleon), negative structural bias ($\beta = -2.0$) remains optimal regardless of w_{self} . Second, on homophilic networks (Cora), performance peaks at unamplified self-loops ($w_{\text{self}} = 1.0\times$, $\beta = 0.0$ at 87.36%) and gradually deteriorates with larger multipliers, confirming that local aggregation is beneficial when neighbors share class labels. On sparse heterophilic networks (Wisconsin), w_{self} and β exhibit strong positive synergy: Scaling w_{self} to 10.0 $\times$ under $\beta = +2.0$ improves validation accuracy from 71.57% to a peak of 84.51%, as heavy self-loops shield ego-node representations from sparse neighborhood noise. In contrast on dense heterophilic networks (Chameleon), negative structural bias ($\beta = -2.0$) achieves its peak at $w_{\text{self}} = 1.0\times$ (68.99%), demonstrating that aggressive low-degree routing sufficiently suppresses cross-class interference without requiring heavy diagonal damping.

TABLE VI. Network depth ablation results (validation accuracy mean $\pm$ Std dev, ten random splits).

Depth	Chameleon (%)	Cora (%)	Wisconsin (%)
1	63.00 $\pm$ 3.32	85.80 $\pm$ 2.87	79.41 $\pm$ 5.96
2	67.13 $\pm$ 1.95	87.11 $\pm$ 2.78	78.04 $\pm$ 5.46
3	65.53 $\pm$ 1.81	86.63 $\pm$ 2.42	72.75 $\pm$ 4.15
4	64.36 $\pm$ 1.90	85.49 $\pm$ 2.74	69.22 $\pm$ 4.30
5	61.60 $\pm$ 1.41	84.74 $\pm$ 3.59	67.06 $\pm$ 5.25

TABLE VII. Average GPU training time.

Dataset	Model	Time (s)
Cora	GAT	1.61
	PDRW-GCN ($\beta = 0.0$)	2.30
	GCN	2.44
Actor	GAT	1.59
	PDRW-GCN ($\beta = 0.0$)	12.33
	GCN	12.31
Texas	GAT	1.47
	PDRW-GCN ($\beta = 2.0$)	0.43
	GCN	0.48
Wisconsin	GAT	1.54
	PDRW-GCN($\beta = 2.0$)	0.47
	GCN	0.51
Cornell	GAT	1.52
	PDRW-GCN ($\beta = 2.0$)	0.45
	GCN	0.43
Chameleon	GAT	1.55
	PDRW-GCN ($\beta = -2.0$)	2.52
	GCN	2.54
Squirrel	GAT	2.52
	PDRW-GCN($\beta = -2.0$)	11.90
	GCN	11.89

Overall, the consistent relative superiority of degree-biased routing across all fixed values of w_{self} confirms that β -driven routing provides an independent, topology-aware performance advantage beyond simple ego-feature retention.

2. Network depth

To investigate the interaction between topology-dependent routing and propagation depth, we varied the number of graph convolutional layers from one to five (Table VI). For both Cora and Chameleon, performance follows the characteristic trajectory associated with graph oversmoothing: Accuracy initially improves at shallow depths before progressively degrading as repeated neighborhood aggregation homogenizes node representations.

The Wisconsin dataset exhibits a different behavior, achieving peak performance with a single propagation layer. This indicates that even minimal message passing introduces substantial semantic interference in sparse heterophilic topologies with high cross-class connectivity. Although the proposed popularity-driven routing operator delays oversmoothing by regulating hub accessibility, the degradation observed at higher layer counts demonstrates that local topological routing alone cannot fully eliminate the global homogenization inherent to deep message-passing architectures. The depth-dependent trends are reported in Table VI.

G. Computational complexity

Theoretical complexity bounds require empirical validation. To assess real-world efficiency, we tracked the training wall-clock durations of standard GCN [1], GAT [10], and the proposed PDRW-GCN framework in a CUDA-enabled environment via Google Colab. Table VII reports average

wall-clock run times (in seconds) alongside classification accuracy metrics across five independent runs.

PDRW-GCN maintains the $O(|E|)$ linear computational complexity and memory footprint of standard GCN message passing while avoiding the computational overhead associated with attention-based or higher-order heterophily architectures.

GAT dynamically computes attention coefficients across every edge during each forward pass [$O(|E| \cdot d)$ per epoch]. In contrast, PDRW-GCN relies on static, degree-based precomputed edge reweighting, incurring zero run-time overhead during training.

Minor empirical run-time variations between standard GCN and PDRW-GCN stem primarily from PyTorch geometric GPU kernel execution and sparse tensor indexing rather than algorithmic complexity. PDRW-GCN thus offers a computationally efficient, nonparametric alternative to learned attention for regulating topological information flow in complex networks.

V. CONCLUSION AND FUTURE WORK

In this work, we presented the PDRW-GCN, a message-passing architecture that reformulates graph convolution as a degree-biased diffusion process. By modulating neighborhood aggregation through a structural bias parameter β , the framework dynamically regulates how information flows through topological hubs across different graph structures. Our experimental evaluations reveal a distinct topology-dependent transport transition: Sparse heterophilic graphs perform best under positive diffusion bias, where structural hubs serve as essential stabilizing transport centers, whereas dense heterophilic networks benefit from negative bias, which suppresses noisy superhubs and prevents feature overmixing. In contrast, homophilic and non-scale-free graphs exhibit relatively low sensitivity to structural bias, demonstrating that the efficacy of degree-biased routing relies on heavy-tailed degree heterogeneity. Importantly, PDRW-GCN achieves these structural adaptations while maintaining the linear run-time complexity of standard GCNs, circumventing the substantial computational overhead of attention-based routing mechanisms. The empirical evidence underlying these topology-dependent diffusion transitions is detailed across Tables III–VI, as well as the bias sweeps and latent-space visualizations in Figs. 1–4.

The current framework has a few key limitations. It remains specifically coupled to scale-free transport structure and thus provides limited benefit on graphs lacking heavy-tailed degree organization. Furthermore, the present formulation relies on a statically selected bias parameter β rather than localized or dynamically learned transport rates.

Future work will investigate addressing these constraints by introducing abrasive or layer-adaptive bias parameters that infer optimal routing preferences end-to-end during training. Theoretical efforts may examine the spectral properties and feature localization dynamics of degree-biased transport operators to establish tighter bounds on oversmoothing and oversquashing. Additionally, extending this diffusion framework to inductive architectures such as GraphSAGE, as well as to temporal and multilayer graph structures, remains a promising direction. Exploring nonlinear transport dynamics derived from generalized degree-biased random walks may offer even finer control over structural information routing. Finally, evaluating PDRW-GCN across controlled synthetic scale-free graph ensembles like Barabási-Albert and Watts-Strogatz networks will help isolate the precise structural metrics that govern optimal transport behavior.

DATA AVAILABILITY

All datasets used in this study are publicly available benchmark graph datasets obtained from previously published sources. Processed data and implementation details are available from the corresponding author upon reasonable request.

- [1] T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, in *5th International Conference on Learning Representations (ICLR 2017), Toulon, France, April 24-26, 2017, Conference Track Proceedings* (OpenReview.net, 2017).
- [2] W. L. Hamilton, Z. Ying, and J. Leskovec, Inductive representation learning on large graphs, in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, Dec. 4-9, 2017* (Curran Associates, Inc., 2017), pp. 1024–1034.
- [3] M. McPherson, L. Smith-Lovin, and J. M. Cook, Birds of a feather: Homophily in social networks, *Annu. Rev. Sociol.* **27**, 415 (2001).
- [4] J. Zhu, Y. Yan, L. Zhao, M. Heimann, L. Akoglu, and D. Koutra, Beyond homophily in graph neural networks: Current limitations and effective designs, *Adv. Neural Inf. Process. Syst.* **33**, 7793 (2020).
- [5] O. Platonov, D. Kuznedelev, M. Diskin, A. Babenko, and L. Prokhorenkova, A critical look at the evaluation of GNNs under heterophily: Are we really making progress? in *International Conference on Learning Representations (ICLR 2023), Kigali, Rwanda, May 15, 2023* (OpenReview.net, 2023).
- [6] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286**, 509 (1999).
- [7] M. E. J. Newman, The structure and function of complex networks, *SIAM Rev.* **45**, 167 (2003).
- [8] Q. Li, Z. Han, and X.-M. Wu, Deeper insights into graph convolutional networks for semi-supervised learning, in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, USA, Feb. 27, 2018* (AAAI Press, Palo Alto, CA, 2018), Vol. 32, pp. 3538–3545.
- [9] K. Oono and T. Suzuki, Graph neural networks exponentially lose expressive power for node classification, in *International Conference on Learning Representations (ICLR 2020), Addis Ababa, Ethiopia, April 26–30, 2020* (OpenReview.net, 2020).
- [10] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, Graph attention networks, in *6th International Conference on Learning Representations (ICLR 2018), Vancouver, BC, Canada, April 30-May 3, 2018, Conference Track Proceedings* (OpenReview.net, 2018).
- [11] H. Pei, B. Wei, K. C.-C. Chang, Y. Lei, and B. Yang, Geom-GCN: Geometric graph convolutional networks, in *International Conference on Learning Representations (ICLR 2020), Addis Ababa, Ethiopia, April 26–30, 2020* (OpenReview.net, 2020).
- [12] F. Xia, L. Wang, T. Tang, X. Chen, X. Kong, G. Oatley, and I. King, CenGCN: Centralized convolutional networks with vertex imbalance for scale-free graphs, *IEEE Trans. Knowl. Data Eng.* **35**, 4555 (2022).
- [13] D. Bo, X. Wang, C. Shi, and H. Shen, Beyond low-frequency information in graph convolutional networks, in *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)* (AAAI Press, Palo Alto, CA, 2021), Vol. 35, pp. 3950–3957.
- [14] E. Chien, J. Peng, P. Li, and O. Milenkovic, Adaptive universal generalized PageRank graph neural network, in *International Conference on Learning Representations (ICLR 2021), Virtual Conference, May 3–7, 2021* (OpenReview.net, 2021).
- [15] G. Nikolentzos and M. Vazirgiannis, Random walk graph neural networks, *Adv. Neural Inf. Process. Syst.* **33**, 16211 (2020).
- [16] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, Search in power-law networks, *Phys. Rev. E* **64**, 046135 (2001).
- [17] A. Fronczak and P. Fronczak, Biased random walks on complex networks: The role of local navigation rules, *Phys. Rev. E* **80**, 016107 (2009).
- [18] J. D. Noh and H. Rieger, Random walks on complex networks, *Phys. Rev. Lett.* **92**, 118701 (2004).
- [19] J. Gómez-Gardenes and V. Latora, Entropy rate of diffusion processes on complex networks, *Phys. Rev. E* **78**, 065102(R) (2008).
- [20] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguná, and D. Krioukov, Popularity versus similarity in growing networks, *Nature (London)* **489**, 537 (2012).
- [21] A. P. Riascos and J. L. Mateos, Random walks on weighted networks: A survey of local and non-local dynamics, *J. Complex Networks* **9**, cnab032 (2021).
- [22] S. Santini, Popularity-driven random walks on a class of scale-free graphs, *Phys. Rev. E* **111**, 054303 (2025).
- [23] K. Singh, N. RV, V. K. Chandrasekar, and D. V. Senthilkumar, Generalized degree-biased random walk on scale-free networks, *Phys. Rev. E* **112**, 064321 (2025).

- [24] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad, Collective classification in network data, *AI Mag.* **29**, 93 (2008).
- [25] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery, Learning to extract symbolic knowledge from the World Wide Web, in *Proceedings of the Fifteenth National Conference on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference (AAAI-98/IAAI-98)*, Madison, WI, USA, July 26–30, 1998 (AAAI Press, 1998), pp. 509–516.
- [26] B. Rozemberczki, C. Allen, and R. Sarkar, Multi-scale attributed node embedding, *J. Complex Networks* **9**, cnab014 (2021).
- [27] L. Tang and H. Liu, Relational learning via latent social dimensions, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, Paris, France, June 28 July 1, 2009 (Association for Computing Machinery, New York, NY, 2009), pp. 817–826.
- [28] M. Fey and J. E. Lenssen, Fast graph representation learning with PyTorch geometric, in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, New Orleans, LA, USA (2019).
- [29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, PyTorch: An imperative style, high-performance deep learning library, *Adv. Neural Inf. Process. Syst.* **32**, 8026 (2019).
- [30] L. Van der Maaten and G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* **9**, 2579 (2008).