WHEN DO DISTANT DEPENDENCIES MATTER? DIAGNOSTICS FOR LONG-RANGE PROPAGATION IN GNNS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph Neural Networks (GNNs) propagate information locally through message passing. While local propagation is often sufficient for short-range tasks, performance can degrade when distant interactions are required. In this paper, we introduce a diagnostic metric that quantifies the role of long-range propagation. The metric is derived from margin-aligned sensitivities, providing an interpretable measure of the dominance of one-hop neighbors in margin-relevant influence. Using this diagnostic, we show that the need for long-range propagation is dataset- and architecture-dependent, rather than universal. We further demonstrate that this diagnostic metric is predictable from well-studied graph-theoretic measures, aligning with the assumptions of rewiring-based approaches. Finally, we show how the diagnostic can be leveraged during training: we design an additional layer that selectively incorporates sensitivity to long-range dependencies and can be applied to any standard GNN backbone. Experiments on both node- and graph-level benchmarks demonstrate consistent gains over rewiring-based methods, without altering the original graph topology.

1 INTRODUCTION

Graph Neural Networks (GNNs) (Goller & Kuchler, 1996; Scarselli et al., 2008; Bruna et al., 2014) learn from relational structure by propagating and aggregating messages (Gilmer et al., 2017), achieving strong performance across chemistry, social networks, recommendation, and spatiotemporal forecasting (Zhou et al., 2020; Duval et al., 2023; Sharma et al., 2024; Wu et al., 2022; Castro-Correa et al., 2024). When task-relevant evidence lies many hops away, however, performance can degrade. A leading explanation is over-squashing (Alon & Yahav, 2021; Topping et al., 2022; Di Giovanni et al., 2023): signals routed through sparse or bottlenecked regions are compressed into fixed-size embeddings, diluting information.

Prior work analyzes over-squashing from two angles (Arnaiz-Rodriguez & Errica, 2025). The structural view Alon & Yahav (2021); Topping et al. (2022) attributes failure to graph bottlenecks, e.g., negative curvature (Topping et al., 2022), low effective resistance (Black et al., 2023), and small spectral gaps (Karhadkar et al., 2023) and typically intervenes by rewiring the graph structure (Attali et al., 2024b; Akansha, 2025). The other angle is the computational (i.e., model) bottlenecks perspective, where the limitation stems from message passing itself: finite-depth, locality aggregations restrict the receptive field and progressively compress signals, so information from long-distance nodes is hard to both reach and preserve (Di Giovanni et al., 2023; Arnaiz-Rodriguez & Errica, 2025). These perspectives are complementary, yet it often remains unclear whether accuracy gains after rewiring truly stem from alleviating over-squashing, or when increasing long-range influence is desirable for the task at hand.

We take a task-aligned view and argue that dependence on long-range information is inherently task- and node-specific. Rather than universally amplifying long-range influence, interventions should be adaptive. To ground this claim, we introduce a margin-aligned, Jacobian-based sensitivity index that quantifies, for a trained GNN, how a node's (or graph's) true classification margin responds to one-hop versus multi-hop perturbations (§3). The resulting long-range capture index $p_u \in [0, 1]$ (with graph-level aggregate ρ_G) is a bounded, interpretable "share-at-one-hop" measure that directly traces computational limitations of message passing.

Using this diagnostic, we show that GNNs implicitly operate in local and non-local regimes: some nodes behave as if decisions are locally controlled (high p_u), others as if non-local inputs dominate (low p_u). Importantly, the model's performance co-varies with this sensitivity in a task-dependent manner: across benchmarks and backbones, the margin–sensitivity correlation is approximately linear, but its sign and magnitude vary with dataset and architecture. In some settings, reducing the one-hop share helps; in others, preserving locality is beneficial. This organization, along a single sensitivity axis, explains when long-range propagation aids or harms prediction (§3).

We then uncover a bridge to graph structure: the true sensitivity defined via the margin can be predicted from topology alone (§4.1). A sparse linear model (Lasso) on structural indicators widely used in structural rewiring (e.g., curvature and effective resistance) yields accurate, structure-only proxies for p_u at node and graph levels. This link connects structural accounts of over-squashing to model-level behavior and enables label-free estimation at test time.

Finally, we convert these insights into a minimal intervention at readout. We introduce FLAN (§4.2), a rewiring-free, lightweight long-range layer that conditions the classifier on the structure-predicted proxy $\hat{p}_u$. The layer applies a small translation and a one-parameter diagonal reweighting of the encoder representation, effectively letting the readout adapt across local vs. non-local regimes while keeping the encoder and topology unchanged. Empirically, this plug-in improves accuracy across GNN backbones and datasets (§5), offering a simple and time-efficient alternative to graph rewiring.

The main contributions of this paper are summarized as follows:

1. We introduce a task-aligned, Jacobian-based diagnostic of long-range sensitivity at node (p_u) and graph (ρ_G) scales.
2. We demonstrate that this diagnostic is accurately predicted from graph structure via a sparse structural model, linking structural bottlenecks to trained model sensitivity.
3. We provide cross-dataset/backbone evidence that margins vary monotonically along the sensitivity axis, with task-dependent sign.
4. Finally, we design FLAN, a rewiring-free, parameter-efficient conditioning layer that leverages the predicted sensitivity $\hat{p}_u$ to improve performance without changing the graph or increasing depth.

Our study contributes to a unified understanding of over-squashing: structural features forecast a trained model's long-range *sensitivity*; errors organize along this sensitivity axis; and an adaptive, low-capacity correction exploits this organization to deliver consistent gains (Arnaiz-Rodriguez & Errica, 2025; Bechler-Speicher et al., 2025).

Reproducibility. The source code to reproduce our experiments is available[1].

2 Background and Related Work

We start by introducing notations used throughout this paper. Let $G = (V, E$ be a simple, undirected, unweighted graph with node-feature matrix $\mathbf{H} \in \mathbb{R}^{|V|\times d}$. Let $\mathbf{A} \in \{0,1\}^{|V|\times|V|}$ be its adjacency matrix, $\mathbf{D} = \mathrm{diag}(d_u)_{u\in V}$ the degree matrix, $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$ the transition matrix and the normalized Laplacian is $\mathbf{L}_{\mathrm{norm}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. For $u \in V$, we denote its neighborhood by $\mathcal{N}(u) = \{\, v \in V : (u, v) \in E \,\}$.

Message passing in GNNs. GNNs are built upon the message passing mechanism, in which node representations are refined through local interactions (Gilmer et al., 2017). At each layer, a node aggregates information from its neighbors using a permutation-invariant function, followed by a learnable transformation. Formally, for a node $i \in \mathcal{V}$, its representation at layer $k+1$ is defined as:

$$\mathbf{h}_i^{(k+1)} = \phi\left(\mathbf{h}_i^{(k)}, \bigoplus_{j\in\mathcal{N}(i)} \psi(\mathbf{h}_j^{(k)})\right),$$

[1] https://anonymous.4open.science/r/FLAN_ICLR_2026-3E65

where $\mathbf{h}_i^{(k)}$ denotes the representation of node i at layer k, ψ the message function, and ϕ the update function. The operator $\bigoplus$ denotes a permutation-invariant aggregation function such as summation, mean, or maximum. This iterative procedure allows GNNs to integrate both feature and structural information from local neighborhoods. Message passing is effective when task-relevant information is local and can be aggregated within only a few hops, which is typically the case in homophilic graphs (Zhu et al., 2021). For long-range dependencies, communication across a distance d requires $\mathcal{O}(d)$ message-passing layers (Barceló et al., 2020). Increasing the depth in this way amplifies over-squashing (Di Giovanni et al., 2023; Akansha, 2025) and over-smoothing (Rusch et al., 2023; Giraldo et al., 2023).

Over-squashing, long-range interactions, and graph rewiring. Over-squashing occurs when information from exponentially large neighborhoods must be compressed into fixed-size node embeddings within a limited number of message-passing layers (Alon & Yahav, 2021; Topping et al., 2022). As the receptive field expands with depth, the aggregation function is forced to encode ever larger amounts of information into a bounded representation, creating a bottleneck that severely limits the ability of GNNs to capture long-range dependencies, particularly in graphs with sparse connectivity or complex topology.

Graph rewiring addresses over-squashing and long-range dependencies by modifying the input topology of a GNN, alleviating structural bottlenecks that hinder the propagation of information across distant nodes. Early work focuses on curvature-based rewiring, adding edges around regions with highly negative discrete curvature that indicate bottlenecks (Topping et al., 2022; Giraldo et al., 2023; Nguyen et al., 2023; Fesser & Weber, 2023). Because discrete curvature measures are inherently local (Forman, 2003; Ollivier, 2007; Samal et al., 2018), subsequent approaches have targeted more global signals, either increasing the spectral gap to improve connectivity and mixing (Banerjee et al., 2022; Karhadkar et al., 2023) or minimizing effective resistance, which models the difficulty of information transmission between node pairs (Black et al., 2023).

More recently, a complementary line of work incorporates node features into the rewiring techniques. For example, Delaunay-based rewiring reconstructs the graph by performing a Delaunay triangulation in feature space, thereby removing edges that exhibit extreme discrete curvature (Attali et al., 2024a; 2025). Other approaches jointly modify the topology and the initial node features to maximize the spectral alignment between the feature signal and the structural information (Linkerhägner et al., 2025). Finally, intra-community rewiring guided by the cosine similarity of node features has been proposed to densify connections among similar nodes while preserving community-level structure (Rubio-Madrigal et al., 2025).

One can distinguish between different types of bottlenecks. Structural bottlenecks arise from the graph's topology (narrow cuts, hubs, or low expansion) that restrict information flow regardless of the model. Computational bottlenecks stem from the message-passing computation itself: even on favorable graphs, signals and gradients from distant nodes attenuate through repeated local updates. Most existing metrics target structural limits; far fewer directly capture the computational one. The computational bottleneck is often studied via Jacobians : Topping et al. (2022); Di Giovanni et al. (2023) show that node-to-node sensitivity decays exponentially with graph distance, explaining the difficulty of propagating long-range information in GNNs.

3 GNN Performance and Long-range Dependencies

In this section, we extend the study of long-range effects and over-squashing by grounding the analysis in the model's Jacobian (Topping et al., 2022; Di Giovanni et al., 2023; Giovanni et al., 2024). Rather than focusing on pairwise dependencies between individual nodes, we directly quantify both the distance (in graph terms) and the amount of task-relevant information that a node's representation can capture in a classification task. Concretely, we aggregate margin-aligned Jacobian sensitivities into a one-hop dominance measure, quantifying how much of the margin-relevant signal is captured locally rather than over longer ranges. We then examine how this long-range signal relates to architectural performance. Importantly, instead of relying solely on accuracy, we evaluate with the classification margin, which provides a finer view of confidence and decision robustness. This margin-aligned perspective allows us to connect distance-structured sensitivity to accuracy gains offering a clear diagnostic of when and how architectures benefit from long-range information. Below, we elaborate on the different steps.

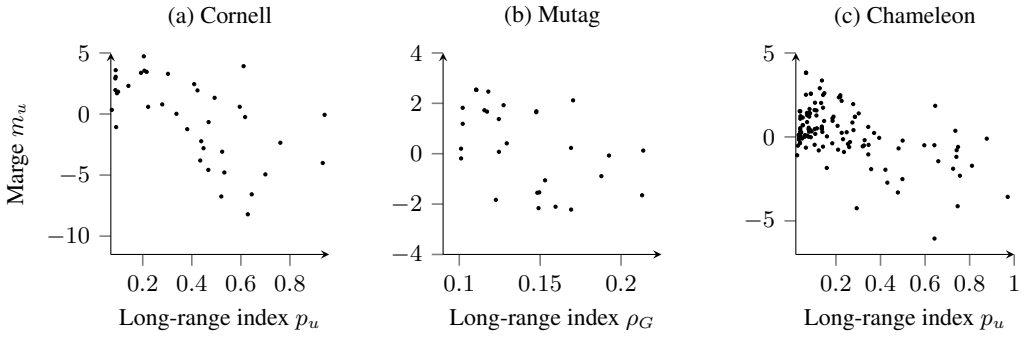

Figure 1: Correlation between classification margin and the long-range capture index $\bar{\rho}_u$: (a) CORNELL (node classification), (b) MUTAG (graph classification), (c) CHAMELEON (node classification). Higher $\bar{\rho}_u$ values (dominant 1-hop contribution) tend to coincide with smaller margins when long-range evidence is required.

Task-aware node margin. For node classification, let $\mathbf{z}_u \in \mathbb{R}^C$ be the logits predicted for node $u \in V$ with ground-truth label $y_u \in \{1, \ldots, C\}$. The node-level margin is defined as

$$m_u \;=\; \mathbf{z}_u[y_u] \;-\; \max_{c \neq y_u} \mathbf{z}_u[c]. \tag{1}$$

The margin is directly aligned with the downstream task: $m_u > 0$ indicates correct classification; larger values reflect a larger separation from the closest competing class. For graph classification, we similarly define a graph-level margin m_G.

Label-aware sensitivity. To attribute the classification margin to input features, we compute the magnitude of the first-order effect:

$$J^u_{s,g} \;:=\; \Big|\frac{\partial m_u}{\partial \mathbf{H}^{(0)}_{s,g}}\Big|, \tag{2}$$

where $s \in V$ indexes a source node and $g \in \{1, \ldots, F\}$ a feature dimension. Intuitively, $J^u_{s,g}$ measures how much the classification margin of node u changes in response to a small change in feature g of source node s.

Distance-binned aggregation. Having computed the label-aware sensitivities, we next aggregate them according to graph distance from a reference node u:

$$S_{u,g}(k) \;=\; \sum_{s:\, \mathbf{D}(s,u)=k} J^u_{s,g}, \qquad k = 0, 1, 2, \ldots, \tag{3}$$

with $\mathbf{D}(\cdot, \cdot)$ the number-of-hops on the input graph. This yields a distance-resolved profile of label-aware influence; in message passing GNNs, contributions beyond the network depth are typically negligible, but we retain the full histogram for completeness.

Long-range capture index. We quantify the fraction captured only by the one-hop neighborhood; for a node u we define:

$$\rho_{u,g} \;=\; \frac{S_{u,g}(1)}{\sum_{k \geq 1} S_{u,g}(k)} \;\in [0, 1]. \tag{4}$$

Normalizing by $\sum_{k \geq 1}$ makes $\rho_{u,g}$ scale-invariant to global rescalings of gradients. We obtain a node- and a graph-level score by averaging over features as follows:

Node-level index

$$p_u \;=\; \frac{1}{F} \sum_{g=1}^{F} \rho_{u,g} \in [0, 1]. \tag{5}$$

Graph-level index

$$\rho_G \;=\; \frac{1}{|V|\, F} \sum_{u \in V} \sum_{g=1}^{F} \rho_{u,g} \;\in [0, 1]. \tag{6}$$

Larger p_u (and ρ_G) indicates that margin-relevant influence is disproportionately concentrated at distance 1, indicating limited long-range transmission to u. Equivalently, this long-range diagnostic

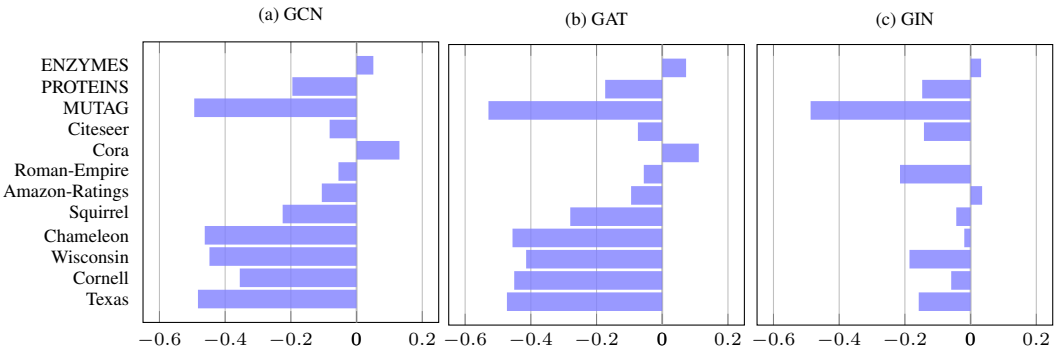

Figure 2: Correlation (mean over 20 runs) between the classification margin and the long-range capture index ρ_u for three backbones (GCN, GAT, GIN) across node and graph level benchmarks. Negative values indicate that performance increases as the 1-hop share decreases, i.e., when long-range propagation becomes more informative.

index can be interpreted as a one-hop dominance score for node u: it summarizes how much of the margin-aligned sensitivity that reaches u is already captured in its immediate neighborhood as opposed to arriving from longer ranges.

How does the long-range capture index relate to the classification task? To analyze GNN's behavior on a given graph dataset, we study the correlation between the classification margin m_u (Eq. (1)) and the long-range capture index (Eq. (5) and (6)). Figure 1 illustrates the trends on Chameleon, Cornell, and MUTAG. Figure 2 reports the mean correlation over 20 runs for GCN (Kipf & Welling, 2017), GAT (Veličković et al., 2018), and GIN (Xu et al., 2019) across nine node and three graph classification datasets commonly used in graph rewiring experiments (Topping et al., 2022; Giraldo et al., 2023; Attali et al., 2024a; Karhadkar et al., 2023; Nguyen et al., 2023; Liang et al., 2025). Experimental details are provided in Appendix A.1.

Across datasets, the correlation between the classification margin and the long-range capture index is not universal but depends on both the dataset and the GNN backbone. On heterophilic graphs, GCN and GAT exhibit consistently negative correlations, indicating that margins improve as reliance on one-hop information decreases, i.e., long-range capture helps. On homophilic graphs, the correlation is close to zero and slightly positive, indicating that one-hop information is more informative for the task than long-range information, which aligns with the inherent structure of the graph. GCN and GAT exhibit broadly similar behavior on node classification datasets: their diffusion-oriented aggregation induces greater variability in the one-hop share p_u. In contrast, GIN operates in a distinct regime: its sum aggregation followed by an MLP prioritizes strictly local evidence, yielding larger and more tightly concentrated p_u and a reduced reliance on long-range contributions.

4 From Diagnostics to Long-range Intervention

4.1 Decoding Long-range Effects from Graph Topology

To mitigate long-range dependencies, rewiring methods typically rely on structural indicators. In this section, we ask whether topology alone can explain and predict the node-wise long-range capture index, i.e., whether the structural indicators used for rewiring recover p_u or ρ_G. To obtain an interpretable link between graph topology and our diagnostic index, we estimate a sparse linear relation whose coefficients identify the indicators that affect p_u (or ρ_G) along with the sign and magnitude of their effects. To this end, we use four measures that are widely used in graph rewiring methods.

(i) PageRank (Page et al., 1999). PageRank measures how much random walk probability accumulates around a node; high scores identify hub-like vertices. Leveraging this centrality, GNN methods for centrality-aware rewiring and capacity allocation either (i) capture higher-order connectivity (Klicpera et al., 2019), (ii) introduce a highly central virtual master node and connect it to targeted regions (Qian et al., 2024; Southern et al., 2025), or (iii) allocate greater feature capacity

Dataset	GCN	GAT	GIN
Texas	$0.6377_{\pm 0.10}$	$0.3859_{\pm 0.14}$	$0.5127_{\pm 0.10}$
Cornell	$0.7037_{\pm 0.15}$	$0.5161_{\pm 0.18}$	$0.5560_{\pm 0.11}$
Wisconsin	$0.5653_{\pm 0.11}$	$0.4269_{\pm 0.12}$	$0.5453_{\pm 0.09}$
Chameleon	$0.4270_{\pm 0.05}$	$0.3509_{\pm 0.04}$	$0.6769_{\pm 0.28}$
Squirrel	$0.4349_{\pm 0.01}$	$0.3258_{\pm 0.19}$	$0.4349_{\pm 0.01}$
Amazon-Ratings	$0.7897_{\pm 0.01}$	$0.4000_{\pm 0.04}$	$0.8055_{\pm 0.02}$
Roman-empire	$0.6831_{\pm 0.03}$	$0.6551_{\pm 0.03}$	$0.4070_{\pm 0.04}$
Cora	$0.3050_{\pm 0.02}$	$0.3600_{\pm 0.02}$	$0.2737_{\pm 0.03}$
Citeseer	$0.3100_{\pm 0.03}$	$0.3644_{\pm 0.04}$	$0.3377_{\pm 0.04}$
MUTAG	$0.9922_{\pm 0.00}$	$0.9751_{\pm 0.03}$	$0.9867_{\pm 0.01}$
PROTEINS	$0.9560_{\pm 0.01}$	$0.9564_{\pm 0.01}$	$0.9531_{\pm 0.01}$
ENZYMES	$0.7548_{\pm 0.13}$	$0.7548_{\pm 0.13}$	$0.7862_{\pm 0.09}$
IMDB	$0.8340_{\pm 0.01}$	$0.8349_{\pm 0.01}$	$0.7567_{\pm 0.02}$

Table 1: R^2 mean on the test set of Lasso regression using structure indicators to predict the capture index across different backbones and datasets.

to high-PageRank nodes to facilitate information flow (Choi et al., 2024). Formally the PageRank vector $\boldsymbol{\pi} \in \mathbb{R}^{|V|}$ satisfies $\boldsymbol{\pi}^\top = (1-\alpha)\,\mathbf{1}^\top/|V| + \alpha\,\boldsymbol{\pi}^\top\mathbf{P}$, where $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$ is row-stochastic, $\alpha \in (0,1)$ is the teleportation parameter.

(ii) Forman–Ricci edge curvature (Samal et al., 2018). Edges with highly negative curvature typically coincide with structural bottlenecks that intensify over-squashing (Alon & Yahav, 2021; Topping et al., 2022), whereas edges with highly positive curvature promote intra-cluster propagation and can accentuate over-smoothing (Nguyen et al., 2023). These curvature signals motivate curvature-aware rewiring that targets bottlenecks to improve information flow (Topping et al., 2022; Giraldo et al., 2023; Nguyen et al., 2023; Fesser & Weber, 2023; Liu et al., 2023). For an edge $e = (u, v)$, we use the augmented Forman curvature $F(u, v) = 4 - (d_u + d_v) + 3\,t_{uv}$, where t_{uv} is the number of triangles incident to (u, v). Let $q_{0.1}$ and $q_{0.9}$ denote the 10th and 90th percentiles of $\{F(e)\}_{e\in E}$. To obtain node-level indicators, for each node u we count incident edges in the bottom and top deciles: $F_{10}(u) = |\{v \in N(u) : F(u, v) \le q_{0.1}\}|, \quad F_{90} = |\{v \in N(u) : F(u, v) \ge q_{0.9}\}|$. A large $\mathtt{bot}_{0.1}(u)$ signals exposure to strongly negative-curvature (bottleneck) edges, while a large $\mathtt{top}_{0.9}(u)$ characterizes cohesive, intra-cluster ties.

(iii) Mean commute time. Commute time quantifies the difficulty of long-range transmission, large values highlight regions where propagation is inefficient and motivate rewiring to improve long range connectivity (Di Giovanni et al., 2023; Black et al., 2023; Barbero et al., 2024; Sterner et al., 2024; Zhuo et al., 2025). Formally we define the mean commute time as $C_{uv} = 2\,|E|\,R_{uv}$, where R_{uv} is the effective resistance (Chandra et al., 1989) between node u and v. For a node u the mean commute time is defined as $\overline{C}(u) = \frac{1}{|V|-1}\sum_{j\in V\setminus\{u\}} C_{uj}$. Large $\overline{C}(u)$ indicates costly long-range access between u and the rest of the graph (Di Giovanni et al., 2023).

Finally, the node-level structural indicator is the aggregation of four measures:

$$\boldsymbol{s}(u) \;=\; \big[\overline{C}(u),\; \pi(u),\; F_{10}(u),\; F_{90}(u)\big] \in \mathbb{R}^4.$$

Sparse linear model for long-range capture index. Let $\mathbf{S} \in \mathbb{R}^{N\times 4}$ stack $\boldsymbol{s}(u)$ over nodes. We fit a sparse linear predictor of the task-aligned index p_u or ρ_G:

$$(\widehat{\beta}_0, \widehat{\boldsymbol{\beta}}) \;\in\; \arg\min_{\beta_0,\boldsymbol{\beta}} \frac{1}{2|\mathcal{I}_{\text{train}}|} \sum_{u\in\mathcal{I}_{\text{train}}} \big(p_u - \beta_0 - \mathbf{S}_u^\top\boldsymbol{\beta}\big)^2 \;+\; \lambda\|\boldsymbol{\beta}\|_1, \tag{7}$$

with λ chosen by K-fold cross-validation on training nodes. We report the test $R^2(\hat{p}_u, p_u)$ in Table 1.

Can structure alone predict the long-range capture index? On graph classification, the structure-only proxy closely matches the model-derived p_u. For node classification, the alignment is strongest on heterophilous datasets and attenuates on homophilous ones, where one-hop evidence dominates. These trends hold across GCN, GAT, and GIN, indicating robustness to the backbone.

They further confirm that p_u and ρ_u are largely topology-driven, reflecting the same structural signals that rewiring methods leverage.

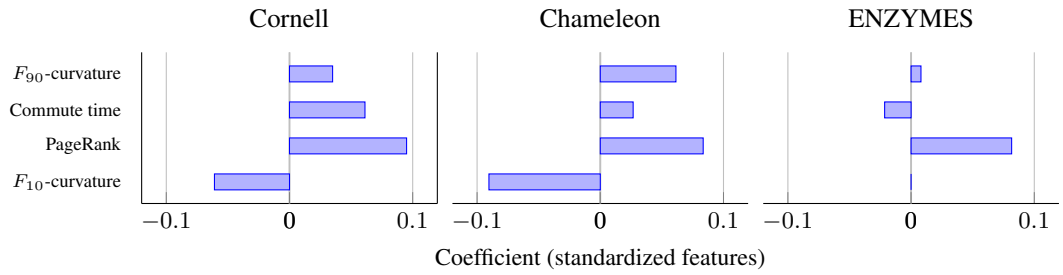

Figure 3: Lasso node indicators influencing the long-range capture index.

Analysis of the Lasso coefficients. Figure 3 reports the Lasso coefficients on three datasets. The coefficients vary across datasets, indicating that different structural indicators modulate the long-range capture index. For instance, a lower mean commute time corresponds to a slight increase in p_u on Cornell, whereas it correlates negatively with ρ_G on ENZYMES. We also observe that higher PageRank, i.e., greater centrality, typically coincides with a high p_u, suggesting that highly central nodes struggle to capture long-range information; their capacity concentrates on strong one-hop signals and thereby reduces the contribution of distant nodes, in line with Choi et al. (2024).

Topological bottlenecks and long-range propagation. For some datasets we observe that incidence to highly negative Forman–curvature edges is negatively associated with p_u. While such edges are often labeled as bottlenecks, negatively curved edges can also act as bridges linking distinct communities: being incident to one effectively grants a node access to many distant neighbors, which lowers p_u. This result corroborates the observation of Arnaiz-Rodriguez & Errica (2025) that not all bottlenecks are harmful to long-range dependence; some enable controlled long-range reach.

Can structural properties predict the true classification margin? We replaced the sparse linear model of (7) used to predict the long-range index p_u from our four node-level structural indicators with an otherwise identical Lasso that instead regresses the true node margin m_u from the same features. On held-out nodes across datasets and backbones, this topology-only regression of m_u yielded very low $R^2(\approx 0)$, in contrast to the substantially higher R^2 obtained when predicting p_u. This outcome is consistent with our framework: m_u is jointly determined by labels, node features, and the learned encoder, and its association with long-range effects even changes sign across datasets, whereas p_u isolates a one-hop-dominance property that is largely structural and thus predictable from these indicators. In short, topology helps locate where long-range pressure exists, but it cannot by itself reconstruct how confident the model is in a class decision.

4.2 FLAN: A Rewiring-free Long-range Layer

Our analysis shows that node margins vary systematically with p_u (§3); and that p_u is predictable from structure alone (§4.1). A single global linear head must therefore compromise across local vs. non-local regimes. We propose a topology-preserving readout adjustment whose per-node intensity is driven by the measurable diagnostic $\hat{p}_u$, without changing the graph or increasing depth (main results and ablation in §5).

Setup. Let $\Phi_{\boldsymbol{\theta}}$ be a frozen GNN encoder with L message passing layers and let $\mathbf{h}_u^{(L)} = \Phi_{\boldsymbol{\theta}}(\cdot)_u \in \mathbb{R}^d$ be the embedding of node (or graph) u. Let $p_u \in [0,1]$ denote the long-range capture index defined in (5); we estimate it per node using the sparse linear model of (7) over structural indicators, yielding $\hat{p}_u \in [0,1]$.

FLAN Layer. We attach a gating map $g_\phi : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ with parameters $\phi = (\boldsymbol{w}_\gamma, \boldsymbol{w}_\beta)$, $\boldsymbol{w}_\gamma, \boldsymbol{w}_\beta \in \mathbb{R}^d$:

$$\mathbf{z}_u \;=\; \sigma_s\big(\boldsymbol{w}_\gamma\, \hat{p}_u\big) \odot \mathbf{h}_u^{(L)} \;+\; \boldsymbol{w}_\beta\, \hat{p}_u, \tag{8}$$

where $\sigma_s(\cdot)$ is an elementwise sigmoid and $\odot$ is the Hadamard product. The classifier is linear,

$$\text{logits}(u) = \mathbf{W}\mathbf{z}_u + b, \qquad \mathbf{W} \in \mathbb{R}^{C\times d},\ \mathbf{b} \in \mathbb{R}^C. \tag{9}$$

During training, we optimize $(\phi, \mathbf{W}, \mathbf{b})$ with cross-entropy; $\boldsymbol{\theta}$ is kept fixed. Intuitively, FLAN uses the diagnostic signal $\hat{p}_u$ to apply a per-node rescaling of $\mathbf{h}_u^{(L)}$ and a per-node bias shift in logit space.

Geometric view. The additive term $\boldsymbol{w}_\beta \hat{p}_u$ implements a p-dependent *translation* of the decision boundary (a family of parallel hyperplanes indexed by $\hat{p}$). The multiplicative term $\sigma_s(\boldsymbol{w}_\gamma \hat{p}_u) \odot \mathbf{h}_u^{(L)}$ implements a p-dependent *reweighting* of coordinates, effectively tilting the separator. The sensitivity index compresses long-range demand into a single axis that is highly predictive of where the baseline fails. Because the dominant error varies monotonically with p_u, this rank-1 translation plus diagonal reweighting is a minimal intervention that corrects the under-performing p regime.

5 Experiments

To evaluate the effect of the proposed FLAN layer, we evaluate it on node classification tasks spanning both homophilic graphs (Sen et al., 2008) and heterophilic graphs (Rozemberczki et al., 2021; Tang et al., 2009), as well as on graph classification benchmarks (Morris et al., 2020). The latter are widely adopted in the evaluation of rewiring methods, since their structures are tightly coupled to the downstream task and require the propagation of long-range dependencies (Karhadkar et al., 2023).

Baseline models. We compare FLAN to seven state-of-the-art rewiring techniques: the curvature-based methods SDRF (Topping et al., 2022) and BORF (Nguyen et al., 2023); the spectral rewiring method FoSR (Karhadkar et al., 2023); the resistance-based approach GTR (Black et al., 2023); LASER (Barbero et al., 2024) a Random Walk Rewiring Based method; DR (Attali et al., 2024a) leverages node features to perform Delaunay triangulation-based rewiring; GOKU (Liang et al., 2025), two-stage densify–then-sparsify rewiring that preserves spectral properties and improves long-range information flow.

Experimental setup. We follow the evaluation protocol of (Liang et al., 2025): GNN hyperparameters are fixed across methods (learning rate $1\mathrm{e}{-3}$, hidden dimension 64, four layers), while rewiring hyperparameters are tuned per method. Baseline results are directly reported from (Liang et al., 2025).

Results. Table 2 reports the results of and graph classification tasks across different GNN backbones. Overall, without altering the input topology, FLAN improves backbone GNN performance by more than 12% on average, and it outperforms recent rewiring baselines. On graph classification, it outperforms all rewiring methods with both GCN and GIN backbones; this is consistent with the higher and more stable R^2 of the structure-only proxy for $\hat{p}_G$, which makes the scalar conditioning particularly effective at the graph level. On node classification, the FLAN layer is competitive but not always state-of-the-art on small heterophilic datasets, where the correlation between margin and p_u (and the corresponding R^2) exhibits high variance, making gains less stable. We also observe benefits on homophilous datasets, where one-hop evidence dominates and the layer acts conservatively rather than over-correcting. Figure 4 shows that the gains arise not from added capacity, but from the long-range signal encoded by the predicted long-range index.

Time comparison. In Appendix 3, we compare FLAN's preprocessing runtime against graph–rewiring baselines. The reported times include (i) Jacobian–margin evaluation, (ii) computation of structural indicators, and (iii) Lasso fitting for $\hat{p}_G$ and $\widehat{m}_G$. On average across datasets, our method is 10^1–$10^3\times$ faster than curvature-based rewiring (Topping et al., 2022; Nguyen et al., 2023), spectral-gap-based rewiring (Karhadkar et al., 2023), and resistance-based rewiring (GTR) (Black et al., 2023).

Ablation studies. To confirm that improvements are diagnostic-driven, Figure 4 compares $\hat{p}$-conditioning to shuffled $\hat{p}$ (permuted across graphs) and to a margin-conditioned scalar (conditioning on margin rather than $\hat{p}$). Only FLAN yields significant, stable gains over the backbone GNN, supporting that the benefits arise from the structure-predicted index rather than added capacity or margin tuning.

(a) Node classification (Backbone: GCN)

Method	Cora	Citeseer	Texas	Cornell	Wisconsin	Chameleon
None	$86.7_{\pm 0.3}$	$72.3_{\pm 0.3}$	$44.2_{\pm 1.5}$	$41.5_{\pm 1.8}$	$44.6_{\pm 1.4}$	$59.2_{\pm 0.6}$
SDRF	$86.3_{\pm 0.3}$	$72.6_{\pm 0.3}$	$43.9_{\pm 1.6}$	$42.2_{\pm 1.5}$	$46.2_{\pm 1.2}$	$59.4_{\pm 0.5}$
FOSR	$85.9_{\pm 0.3}$	$72.3_{\pm 0.3}$	$46.0_{\pm 1.2}$	$40.2_{\pm 1.1}$	$48.3_{\pm 1.3}$	$59.3_{\pm 0.6}$
BORF	$\underline{87.5_{\pm 0.2}}$	$\underline{73.8_{\pm 0.2}}$	$49.4_{\pm 1.8}$	$50.8_{\pm 1.1}$	$50.3_{\pm 0.9}$	$61.5_{\pm 0.4}$
DR	$78.4_{\pm 1.2}$	$69.5_{\pm 1.6}$	$\underline{67.8_{\pm 2.5}}$	$\underline{57.8_{\pm 1.9}}$	$\underline{62.8_{\pm 2.1}}$	$58.6_{\pm 0.8}$
GTR	$87.3_{\pm 0.4}$	$72.4_{\pm 0.3}$	$45.9_{\pm 1.9}$	$50.8_{\pm 1.6}$	$46.7_{\pm 1.5}$	$57.6_{\pm 0.8}$
LASER	$86.9_{\pm 1.1}$	$72.6_{\pm 0.6}$	$45.9_{\pm 2.6}$	$42.7_{\pm 2.6}$	$46.0_{\pm 2.6}$	$43.5_{\pm 1.0}$
GOKU	$86.8_{\pm 0.3}$	$73.6_{\pm 0.2}$	$\mathbf{72.4_{\pm 2.2}}$	$\mathbf{69.4_{\pm 2.1}}$	$\mathbf{68.8_{\pm 1.4}}$	$\underline{63.2_{\pm 0.4}}$
FLAN	$\mathbf{88.3_{\pm 0.9}}$	$\mathbf{75.6_{\pm 0.5}}$	$55.6_{\pm 3.0}$	$51.9_{\pm 3.1}$	$54.5_{\pm 2.9}$	$\mathbf{65.1_{\pm 0.6}}$

(b) Graph classification (Backbone: GCN on the left; GIN on the right)

	Backbone: GCN				**Backbone: GIN**			
	ENZYMES	IMDB	MUTAG	PROTEINS	ENZYMES	IMDB	MUTAG	PROTEINS
None	$27.1_{\pm 1.6}$	$49.5_{\pm 1.0}$	$70.3_{\pm 2.1}$	$71.4_{\pm 1.0}$	$33.5_{\pm 1.3}$	$67.7_{\pm 1.4}$	$76.1_{\pm 3.1}$	$69.5_{\pm 1.4}$
SDRF	$26.1_{\pm 1.1}$	$49.1_{\pm 0.9}$	$70.5_{\pm 2.1}$	$71.4_{\pm 0.8}$	$32.4_{\pm 1.3}$	$69.4_{\pm 1.4}$	$\underline{79.5_{\pm 2.6}}$	$71.4_{\pm 0.8}$
FOSR	$27.4_{\pm 1.1}$	$49.6_{\pm 0.8}$	$75.6_{\pm 1.7}$	$72.3_{\pm 0.9}$	$28.8_{\pm 1.0}$	$70.6_{\pm 1.3}$	$74.8_{\pm 1.5}$	$73.7_{\pm 0.8}$
BORF	$24.7_{\pm 1.0}$	$50.1_{\pm 0.9}$	$75.8_{\pm 1.9}$	$71.0_{\pm 0.8}$	$31.4_{\pm 1.5}$	$70.5_{\pm 1.3}$	$78.2_{\pm 1.6}$	$71.9_{\pm 1.3}$
DR	–	$47.0_{\pm 0.7}$	$80.1_{\pm 1.8}$	$72.2_{\pm 0.8}$	–	$64.8_{\pm 0.8}$	$74.5_{\pm 2.8}$	$\underline{74.3_{\pm 0.8}}$
GTR	$27.4_{\pm 1.1}$	$49.5_{\pm 1.0}$	$78.9_{\pm 1.8}$	$\underline{72.4_{\pm 1.2}}$	$28.4_{\pm 1.8}$	$70.1_{\pm 1.2}$	$78.5_{\pm 3.5}$	$73.3_{\pm 0.9}$
LASER	$\underline{27.6_{\pm 1.3}}$	$\underline{50.3_{\pm 1.3}}$	$78.8_{\pm 1.6}$	$71.8_{\pm 1.6}$	$\underline{35.3_{\pm 1.3}}$	$68.6_{\pm 1.2}$	$76.1_{\pm 2.4}$	$72.1_{\pm 0.7}$
GOKU	$27.6_{\pm 1.2}$	$49.8_{\pm 0.7}$	$\underline{81.0_{\pm 2.0}}$	$71.9_{\pm 0.8}$	$33.8_{\pm 1.2}$	$\underline{71.3_{\pm 0.9}}$	$78.4_{\pm 2.5}$	$73.9_{\pm 1.0}$
FLAN	$\mathbf{33.8_{\pm 1.8}}$	$\mathbf{54.8_{\pm 1.6}}$	$\mathbf{81.2_{\pm 2.5}}$	$\mathbf{74.3_{\pm 1.7}}$	$\mathbf{35.8_{\pm 1.9}}$	$\mathbf{72.0_{\pm 1.3}}$	$\mathbf{81.3_{\pm 2.7}}$	$\mathbf{74.2_{\pm 1.7}}$

Table 2: Performance (%) on node and graph classification benchmarks. Best results are in **bold**, second best are underlined. A dash indicates not applicable.

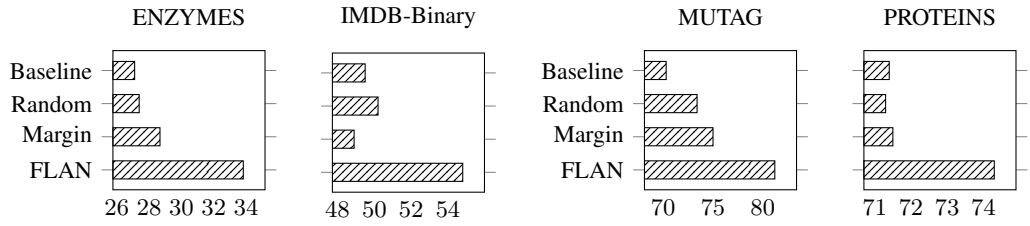

Figure 4: Test accuracy with FLAN vs. (i) random shuffled $\hat{p}$, (ii) margin scalar on graph classification with a GCN as backbone.

In Appendix A.3, we further analyze the mechanism by quantifying both the intervention magnitude (e.g., $\|\mathbf{z} - \mathbf{h}^{(L)}\|$) and the resulting change in task margin in the graph classification task. Graphs with a higher long-range index ρ_G receive stronger corrections from FLAN and achieve larger margin gains, showing that the layer adapts its intervention to the diagnostic's estimate of long-range demand and concentrates changes where they are most needed.

6 Conclusion

We reframed over-squashing as an adaptive, task- and node-specific phenomenon. We (i) defined a margin-aligned sensitivity index for trained GNNs, (ii) showed it is accurately predicted from topology via a sparse structural model, and (iii) found that margins co-vary with this sensitivity with dataset/backbone-dependent sign. Leveraging these findings, we introduced FLAN, a lightweight, rewiring-free readout layer that conditions on a structure-predicted proxy, improving accuracy without changing the graph or deepening the encoder. Our results open promising directions, including targeted rewiring at high-sensitivity nodes. In future work, we will study how this diagnostic can guide and complement graph rewiring methods.

Reproducibility Statement

An anonymized code repository is linked at the end of the Introduction. All datasets, preprocessing steps, fixed splits, hyperparameters, and training/evaluation scripts are specified in the main text and in Appendix A.1, enabling full reproduction of our results.

References

Singh Akansha. Over-squashing in graph neural networks: A comprehensive survey. *Neurocomputing*, pp. 130389, 2025.

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *International Conference on Learning Representations*, 2021.

Adrian Arnaiz-Rodriguez and Federico Errica. Oversmoothing," oversquashing", heterophily, long-range, and more: Demystifying common beliefs in graph machine learning. *arXiv preprint arXiv:2505.15547*, 2025.

Hugo Attali, Davide Buscaldi, and Nathalie Pernelle. Delaunay graph: Addressing over-squashing and over-smoothing using delaunay triangulation. In *Forty-first International Conference on Machine Learning*, 2024a. URL https://openreview.net/forum?id=uyhjKoaIQa.

Hugo Attali, Davide Buscaldi, and Nathalie Pernelle. Rewiring techniques to mitigate oversquashing and oversmoothing in gnns: A survey. *arXiv preprint arXiv:2411.17429*, 2024b.

Hugo Attali, Thomas Papastergiou, Nathalie Pernelle, and Fragkiskos D. Malliaros. Dynamic triangulation-based graph rewiring for graph neural networks. In *ACM International Conference on Information and Knowledge Management, CIKM*, 2025.

Pradeep Kr Banerjee, Kedar Karhadkar, Yu Guang Wang, Uri Alon, and Guido Montúfar. Oversquashing in gnns through the lens of information contraction and graph expansion. In *2022 58th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1–8. IEEE, 2022.

Federico Barbero, Ameya Velingker, Amin Saberi, Michael M. Bronstein, and Francesco Di Giovanni. Locality-aware graph rewiring in GNNs. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=4Ua4hKiAJX.

Pablo Barceló, Egor V Kostylev, Mikael Monet, Jorge Pérez, Juan Reutter, and Juan-Pablo Silva. The logical expressiveness of graph neural networks. In *8th International Conference on Learning Representations (ICLR 2020)*, 2020.

Maya Bechler-Speicher, Ben Finkelshtein, Fabrizio Frasca, Luis Müller, Jan Tönshoff, Antoine Siraudin, Viktor Zaverkin, Michael M Bronstein, Mathias Niepert, Bryan Perozzi, et al. Position: Graph learning will lose relevance due to poor benchmarks. *arXiv preprint arXiv:2502.14546*, 2025.

Mitchell Black, Zhengchao Wan, Amir Nayyeri, and Yusu Wang. Understanding oversquashing in gnns through the lens of effective resistance. In *International Conference on Machine Learning*, pp. 2528–2547. PMLR, 2023.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *ICLR*, 2014.

Jhon A. Castro-Correa, Jhony H. Giraldo, Mohsen Badiey, and Fragkiskos D. Malliaros. Gegenbauer graph neural networks for time-varying signal reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(9):11734–11745, 2024.

Ashok K Chandra, Prabhakar Raghavan, Walter L Ruzzo, and Roman Smolensky. The electrical resistance of a graph captures its commute and cover times. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pp. 574–586, 1989.

Jeongwhan Choi, Sumin Park, Hyowon Wi, Sung-Bae Cho, and Noseong Park. Panda: Expanded width-aware message passing beyond rewiring. *arXiv preprint arXiv:2406.03671*, 2024.

Andreea Deac, Marc Lackenby, and Petar Veličković. Expander graph propagation. In *Learning on Graphs Conference*, pp. 38–1. PMLR, 2022.

Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Lio, and Michael M Bronstein. On over-squashing in message passing neural networks: The impact of width, depth, and topology. In *ICML*, pp. 7865–7885. PMLR, 2023.

Alexandre Duval, Victor Schmidt, Alex Hernández-García, Santiago Miret, Fragkiskos D. Malliaros, Yoshua Bengio, and David Rolnick. FAENet: Frame averaging equivariant GNN for materials modeling. In *International Conference on Machine Learning, ICML*, 2023.

Federico Errica. On class distributions induced by nearest neighbor graphs for node classification of tabular data. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Lukas Fesser and Melanie Weber. Mitigating over-smoothing and over-squashing using augmentations of forman-ricci curvature. In *The Second Learning on Graphs Conference*, 2023.

Robin Forman. Bochner's method for cell complexes and combinatorial ricci curvature. 2003.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. PMLR, 2017.

Francesco Di Giovanni, T. Konstantin Rusch, Michael Bronstein, Andreea Deac, Marc Lackenby, Siddhartha Mishra, and Petar Veličković. How does over-squashing affect the power of GNNs? *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=KJRoQvRWNs.

Jhony H Giraldo, Konstantinos Skianis, Thierry Bouwmans, and Fragkiskos D. Malliaros. On the trade-off between over-smoothing and over-squashing in deep graph neural networks. In *ACM International Conference on Information and Knowledge Management, CIKM*, pp. 566–576, 2023.

Christoph Goller and Andreas Kuchler. Learning task-dependent distributed representations by backpropagation through structure. In *Proceedings of International Conference on Neural Networks (ICNN'96)*, volume 1, pp. 347–352. IEEE, 1996.

Kedar Karhadkar, Pradeep Kr Banerjee, and Guido Montúfar. Fosr: First-order spectral rewiring for addressing oversquashing in gnns. In *International Conference on Learning Representations*, ICLR, 2023.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the International Conference on Learning Representations*, ICLR, 2017.

Johannes Klicpera, Stefan Weißenberger, and Stephan Günnemann. Diffusion improves graph learning. In *Advances in neural information processing systems*, NeurIPS, 2019.

Langzhang Liang, Fanchen Bu, Zixing Song, Zenglin Xu, Shirui Pan, and Kijung Shin. Mitigating over-squashing in graph neural networks by spectrum-preserving sparsification. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=NiMu23k0Ym.

Jonas Linkerhägner, Cheng Shi, and Ivan Dokmanić. Joint graph rewiring and feature denoising via spectral resonance. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=zBbZ2vdLzH.

Yang Liu, Chuan Zhou, Shirui Pan, Jia Wu, Zhao Li, Hongyang Chen, and Peng Zhang. Curvdrop: A ricci curvature based approach to prevent graph neural networks from over-smoothing and over-squashing. In *Proceedings of the ACM Web Conference 2023*, pp. 221–230, 2023.

Christopher Morris, Nils M Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. *arXiv preprint arXiv:2007.08663*, 2020.

Khang Nguyen, Nong Minh Hieu, Vinh Duc Nguyen, Nhat Ho, Stanley Osher, and Tan Minh Nguyen. Revisiting over-smoothing and over-squashing using ollivier-ricci curvature. In *International Conference on Machine Learning*, pp. 25956–25979. PMLR, 2023.

Yann Ollivier. Ricci curvature of metric spaces. *Comptes Rendus Mathematique*, 345(11):643–646, 2007.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. In *Advances in neural information processing systems*, ICLR, 2020.

Chendi Qian, Andrei Manolache, Christopher Morris, and Mathias Niepert. Probabilistic graph rewiring via virtual nodes. *Advances in Neural Information Processing Systems*, 37:28359–28392, 2024.

Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2):cnab014, 2021.

Celia Rubio-Madrigal, Adarsh Jamadandi, and Rebekka Burkholz. Gnns getting comfy: Community and feature similarity guided rewiring. *arXiv preprint arXiv:2502.04891*, 2025.

T Konstantin Rusch, Michael M Bronstein, and Siddhartha Mishra. A survey on oversmoothing in graph neural networks. *arXiv preprint arXiv:2303.10993*, 2023.

Areejit Samal, RP Sreejith, Jiao Gu, Shiping Liu, Emil Saucan, and Jürgen Jost. Comparative analysis of two discretizations of ricci curvature for complex networks. *Scientific reports*, 8(1): 8650, 2018.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model, 2008.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

Kartik Sharma, Yeon-Chang Lee, Sivagami Nambi, Aditya Salian, Shlok Shah, Sang-Wook Kim, and Srijan Kumar. A survey of graph neural networks for social recommender systems. *ACM Computing Surveys*, 56(10):1–34, 2024.

Joshua Southern, Francesco Di Giovanni, Michael M. Bronstein, and Johannes F. Lutzeyer. Understanding virtual nodes: Oversquashing and node heterogeneity. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=NmcOAwRyH5.

Igor Sterner, Shiye Su, and Petar Veličković. Commute-time-optimised graphs for gnns. *arXiv preprint arXiv:2407.08762*, 2024.

Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 807–816, 2009.

Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. *Proceedings of the International Conference on Learning Representations*, 2022.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. ICLR, 2018.

JJ Wilson, Maya Bechler-Speicher, and Petar Veličković. Cayley graph propagation. *arXiv preprint arXiv:2410.03424*, 2024.

Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5):1–37, 2022.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI open*, 2020.

Jiong Zhu, Ryan A Rossi, Anup Rao, Tung Mai, Nedim Lipka, Nesreen K Ahmed, and Danai Koutra. Graph neural networks with heterophily. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11168–11176, 2021.

Wei Zhuo, Han Yu, Guang Tan, and Xiaoxiao Li. Commute graph neural networks. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=29Leye951l.

A Appendix

A.1 Experimental Setup for the Long-range Capture Index

We report here the GNN hyperparameters used to study the correlation between the long-range capture index p_u and the task-aware classification margin across node and graph-level benchmarks in section 3. Our choices follow common evaluation protocols for rewiring methods with standard GNN backbones for both node classification (Pei et al., 2020; Attali et al., 2024a) and graph classification (Errica, 2023; Deac et al., 2022; Karhadkar et al., 2023; Wilson et al., 2024; Liang et al., 2025).

Node classification. We use two layers, dropout 0.5, learning rate 0.005, and early stopping with a patience of 100 epochs. Hidden dimensions are 32 for Texas, Wisconsin, and Cornell; 48 for Squirrel and Chameleon; and 16 for Cora and Citeseer. We adopt a fixed split with 60% of nodes for training, 20% for validation, and 20% for testing.

Graph classification. We use 4 layers, dropout 0.5, learning rate 0.001, hidden dimension 64, and early stopping with a patience of 100 epochs. Datasets are split into 80% training, 10% validation, and 10% testing.

A.2 Time comparison

Model	IMDB-Binary	MUTAG	ENZYMES	PROTEINS
SDRF	5.13257	0.669701	1.71482	3.02873
FoSR	4.54634	4.71567	4.56855	5.04358
BORF	465.408	53.7069	179.573	351.173
GTR	3.39839	1.54127	2.87399	6.49714
PANDA	0.789759	0.246243	0.278594	0.248043
EGP	0.0185697	0.00446963	0.0163198	0.0393348
CGP	0.0211341	0.00438905	0.0166841	0.0348585
FLAN	0.017668	0.013909	0.016429	0.027119

Table 3: Comparison of the preprocessing time to construct each graph rewiring method compared to our FLAN method (in seconds per graph). Table taken from Wilson et al. (2024).

A.3 Ablation Study

In this section, we visualize how graph embeddings change under FLAN as a function of the graph-level long-range index ρ_G. Across datasets (Figures 5, 6, and 7), we observe a clear trend: graphs with higher ρ_G undergo stronger embedding shifts and achieve larger margin gains. This pattern indicates that FLAN does not apply a uniform correction but adapts the intensity of its intervention to the diagnostic's estimate of long-range demand. Importantly, the largest modifications occur precisely for graphs where local evidence dominates and long-range contributions are underrepresented, confirming that the diagnostic successfully identifies the regimes where intervention is most beneficial.

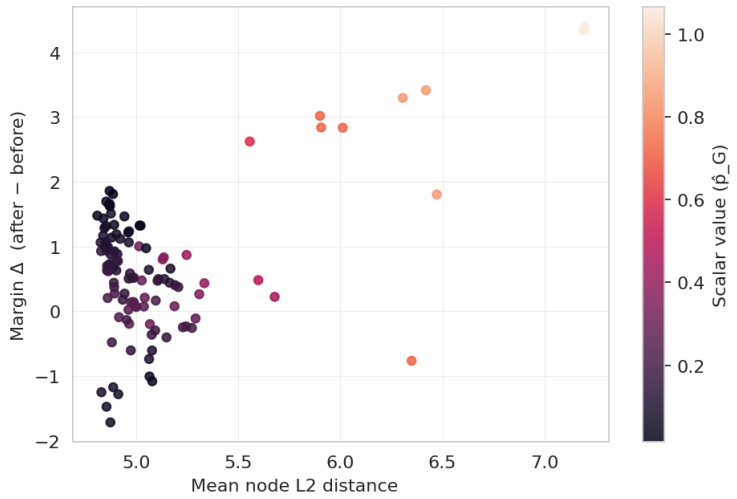

Figure 5: Analysis of FLAN adjustments on PROTEINS: the Euclidean distance $\|\mathbf{z} - \mathbf{h}^{(L)}\|$ (difference between the embedding with FLAN, $\mathbf{z}$, and the backbone embedding without FLAN, $\mathbf{h}^{(L)}$) and the margin gain are plotted against the long-range index ρ_G, showing larger adjustments and stronger improvements as ρ_G increases.

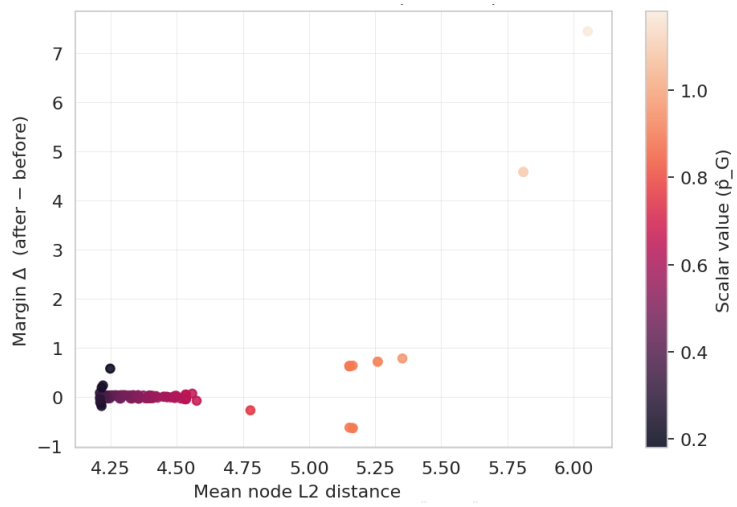

Figure 6: Analysis of FLAN adjustments on IMDB-Binary: the Euclidean distance $\|\mathbf{z} - \mathbf{h}^{(L)}\|$ (difference between the embedding with FLAN, $\mathbf{z}$, and the backbone embedding without FLAN, $\mathbf{h}^{(L)}$) and the margin gain are plotted against the long-range index ρ_G, showing larger adjustments and stronger improvements as ρ_G increases.

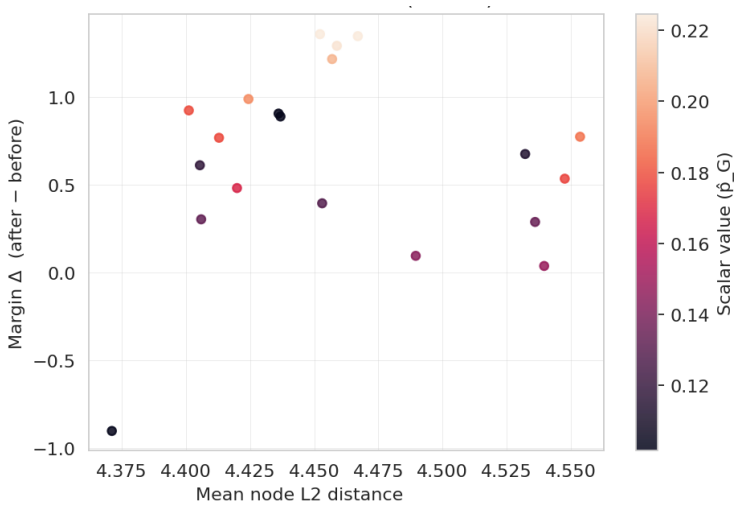

Figure 7: Analysis of FLAN adjustments on MUTAG: the Euclidean distance $\|\mathbf{z} - \mathbf{h}^{(L)}\|$ (difference between the embedding with FLAN, $\mathbf{z}$, and the backbone embedding without FLAN, $\mathbf{h}^{(L)}$) and the margin gain are plotted against the long-range index ρ_G, showing larger adjustments and stronger improvements as ρ_G increases.