Adaptive Self-Interference Control for Large Heterophilic Graphs

Anonymous Author(s)

Abstract

Large models are increasingly used as front-end components for graph learning. However, these semantic interfaces do not solve a structural reliability problem that arises after large-model signals are injected into a graph pipeline: message passing must still decide which neighbor routes should reinforce the target state and which routes may amplify noise, heterophily, or role mismatch. This paper studies this post-semantic structural bottleneck and proposes ***A***daptive Gauge-Invariant ***S***elf-Interference ***C***ontrol for Large Heteroph***i***l***i***c graphs (ASCII). Our method predicts a bounded edge-wise cancellation coefficient from gauge-invariant and label-safe descriptors, suppressing only the target-parallel component of unreliable transported messages while preserving non-parallel evidence for attention. The method is designed as a structural encoder that can be placed after raw attributes, frozen semantic embeddings, or fused large-model features. Theoretical analysis shows that ASCII preserves gauge symmetry, avoids self-parallel energy amplification, and retains sparse message-passing complexity. Experiments on heterophilic benchmarks and large OGB datasets evaluate the structural encoder under matched feature sources and show that adaptive route-wise cancellation improves reliability over fixed self-interference control. The code can be found at https://anonymous.4open.science/r/ASCii-B56F/.

CCS Concepts

• **Computing methodologies** → **Neural networks**; *Supervised learning by classification*; • **Information systems** → *Data mining*; • **Mathematics of computing** → Graph algorithms.

Keywords

Graph machine learning, large-model-augmented graph learning, reliable graph encoding, graph heterophily

ACM Reference Format:

Anonymous Author(s). 2026. Adaptive Self-Interference Control for Large Heterophilic Graphs. In *Proceedings of The 2nd Workshop on Frontiers in Graph Machine Learning for the Large Model Era (GMLLM 2026).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Large models are becoming common front-end modules for graph learning. They can encode text-attributed nodes, summarize relation semantics, produce weak supervision, retrieve graph evidence, and support natural-language reasoning over graph-structured data. This trend shifts part of graph learning from hand-crafted attributes to semantic interfaces produced by large models. Yet it does not eliminate the need for a graph encoder. Once large-model signals are attached to nodes or edges, the model still has to propagate them through an observed graph whose edges may be noisy, heterophilic, directed, or role-changing. Thus, a central question for graph learning in the large-model era is not only how to obtain better semantic features, but also how to prevent unreliable structural routes from amplifying those features in the wrong neighborhood.

Most message-passing GNNs aggregate neighbors with an inductive bias called homophily [16, 24, 31, 32, 38]. This bias is effective on assortative graphs, but it can fail when edges express opposition, role transition, weak affiliation, or label mismatch [29, 30, 39, 45]. A graph with low global homophily may still contain compatible routes, and a graph with high global homophily may still contain harmful routes around a specific node. A single global neighbor-trust rule is too rigid for this setting. Complex-valued and gauge-equivariant graph operators provide a useful language for this problem [5, 9, 10, 19, 25, 42]. A transported complex message can represent direction, polarity, and phase-shifted compatibility in a shared target frame. GESC [8] uses $U(1)$ transport and self-interference cancellation (SIC) to remove the component of an incoming message that is parallel to the current target state. This operation is attractive because it directly targets self-reinforcing buildup.

The limitation is that a fixed cancellation coefficient treats all routes, heads, and layers in the same way. This creates two failure modes. On compatible routes, excessive cancellation can remove useful reinforcement. On noisy, heterophilic, or role-changing routes, insufficient cancellation can leave self-reinforcing errors for attention to amplify. Route reliability can differ across edges around the same node, across heads that represent different compatibility patterns, and across layers where local and higher-order evidence are mixed. A second limitation appears at the output layer. Equivariant propagation is designed so that changing a node-wise complex frame only rotates the hidden state. A classifier based on raw real and imaginary coordinates can still depend on that arbitrary frame. The final prediction can then change under an equivalent representation of the same complex state. This creates a mismatch between the symmetry of propagation and the symmetry of prediction. For graph learning with large models, this distinction should be explicit: semantic features may be strong, but logits should not depend on arbitrary local complex frames.

We propose Adaptive Gauge-Invariant Self-Interference Control (ASCII). For each directed route, attention head, and layer, ASCII predicts a bounded cancellation coefficient from invariant descriptors that do not use validation or test labels. The coefficient only attenuates the self-parallel component of a transported message, so non-parallel information remains available for attention. ASCII also uses a magnitude-Gram readout that is invariant to node-wise $U(1)$ rephasing and still retains projected relative phase relations. The method is agnostic to the feature source: X can be original benchmark attributes, frozen large-model embeddings, or fused semantic features. The contributions are as follows.

- We introduce adaptive SIC, which predicts bounded route-, head-, and layer-specific cancellation coefficients without using validation or test labels.
- We formulate structural reliability as a post-semantic bottleneck for large-model-augmented graph learning and provide several theoretical guarantees.
- The evaluations isolate the structural encoder from the semantics, showing the effectiveness of our method.

2 Related Work

Heterophilic graph learning. Classical GNN algorithms aggregate local neighborhoods using normalized adjacency, attention, or residual propagation [6, 17, 24, 38, 41]. Heterophily-aware methods improve robustness through ego-neighbor separation, higher-order aggregation, geometric neighborhoods, global homophily discovery, adaptive channel mixing, and architecture search [1, 28, 30, 34, 44, 45]. ASCII is complementary to these designs. It estimates how much target-parallel transported signal should be removed before attention aggregates the remaining message.

Spectral, directed, signed, and complex operators. Spectral GNNs use polynomial filters, PageRank coefficients, Bernstein bases, or multi-filter ensembles to control graph frequency responses [7, 11, 14, 18, 40]. Signed and directed GNNs model negative interactions and orientation through signed propagation, magnetic Laplacians, or direction-specific aggregation [12, 19, 21, 25, 35, 42]. Complex embeddings also encode asymmetric relations through rotations and bilinear scores [36, 37]. ASCII uses complex transport as a local message-passing operation and adds bounded route-level interference control. Gauge-equivariant networks formalize the requirement that predictions should not depend on arbitrary local coordinate frames [4, 5, 9, 10]. Reliable graph learning also studies over-smoothing, over-squashing, graph perturbations, denoising, uncertainty, and calibration [2, 22, 23, 27, 33, 43]. ASCII combines these views: it learns an interpretable reliability coefficient for each transported route and aligns the final readout with the gauge symmetry of propagation.

Graph learning in the large-model era. Recent graph learning systems increasingly combine graph neural networks with large models through text encoders, relation descriptions, prompt- or instruction-based supervision, retrieval interfaces, and LLM-assisted augmentation. These systems often improve the semantic input channel, but they still require a graph-side mechanism that decides how information should propagate through the observed structure. ASCII targets this graph-side mechanism. It is not a large model backbone and requires no prompting or instruction tuning. Instead, it provides a reliability-preserving structural encoder that can be placed after different semantic front ends. This position is complementary to LLM-assisted node classification, link prediction, graph generation, and data augmentation: after semantic signals are produced, ASCII controls whether each transported route should reinforce or attenuate the target state. The benefit is most direct in hybrid graph-foundation-model pipelines where strong semantic features coexist with unreliable edges, heterophily, role transitions, or noisy relations.

3 Problem Formulation

Let $G = (V, E, X)$ be an attributed graph with n nodes, feature matrix $X \in \mathbb{R}^{n \times d_x}$, and class labels $c_i \in \{1, \dots, C\}$. During optimization, labels are observed only for training nodes V_{tr}. Validation labels are used only for model selection, and test labels are held out for final reporting. The feature matrix may contain original attributes, frozen large-model embeddings, or fused features. For an undirected graph, each edge is represented by two directed routes $(j \to i)$ in $E^{\to}$. At layer ℓ, node i has a complex state $h_i^{(\ell)} \in \mathbb{C}^d$, and each route uses unit-modulus transport

$$U_{ji}^{(m,\ell)} = \exp(\imath \theta_{ji}^{(m,\ell)}), \qquad \imath^2 = -1. \tag{1}$$

The transport maps the source message into the target frame for head m. Under a local $U(1)$ gauge transformation,

$$h_i^{(\ell)} \mapsto e^{\imath \psi_i} h_i^{(\ell)}, \qquad U_{ji}^{(m,\ell)} \mapsto e^{\imath(\psi_i - \psi_j)} U_{ji}^{(m,\ell)}. \tag{2}$$

A propagation rule is gauge-equivariant when $h_i^{(\ell+1)}$ transforms by the same node-wise phase. A classifier is gauge-invariant when logits are unchanged. This distinction matters because the complex phase is a local frame choice, not a label signal. We focus on semi-supervised node classification under the same train, validation, and test splits. Descriptors used by the adaptive module must be label-safe: they can depend on graph structure, features, frozen unary predictions, layer index, and current hidden states, but they cannot use validation or test labels. Any unary belief u_i is trained only on V_{tr} and then frozen before it is used in descriptors. This restriction is important in transductive node classification because descriptors are computed for routes incident to validation and test nodes. ASCII may use validation labels for hyperparameter selection, but validation and test labels do not enter $z_{ji}^{(m,\ell)}$, transport phases, gates, or unary belief training.

4 Method

As shown in Figure 1, ASCII has four steps. First, real node features are lifted into complex states. Second, each source message is transported into the target frame. Third, a route-wise invariant descriptor predicts how much self-parallel signal to cancel. Fourth, the model aggregates gated messages and applies an invariant readout. An important thing is that cancellation acts before attention, while non-parallel information remains available for aggregation.

4.1 Complex Transport

ASCII first lifts real features into the complex domain,

$$h_i^{(0)} = W_R x_i + \imath W_I x_i. \tag{3}$$

For head m at layer ℓ, the transported source message is

$$\widetilde{h}_{j \to i}^{(m,\ell)} = U_{ji}^{(m,\ell)} W_m^{(\ell)} h_j^{(\ell)}. \tag{4}$$

The phase can be produced by a shared edge encoder from endpoint features, degree features, and edge attributes. If edge attributes are absent, the encoder uses endpoint and degree features. For small transductive benchmarks, the phase can also be stored as a learnable route parameter. In both parameterizations, phases are optimized through the training objective and do not use validation or test labels. Unit-modulus transport changes the target-frame phase but preserves message norm, so later attenuation comes only from the adaptive cancellation and gates.

4.2 Adaptive Self-Interference Cancellation

The goal of SIC is to suppress self-reinforcing buildup without discarding all neighbor information. For target node i, define the Tikhonov-regularized self-alignment operator

$$\Pi_\epsilon(h_i^{(\ell)}) = \frac{h_i^{(\ell)} (h_i^{(\ell)})^H}{\|h_i^{(\ell)}\|_2^2 + \epsilon}. \tag{5}$$

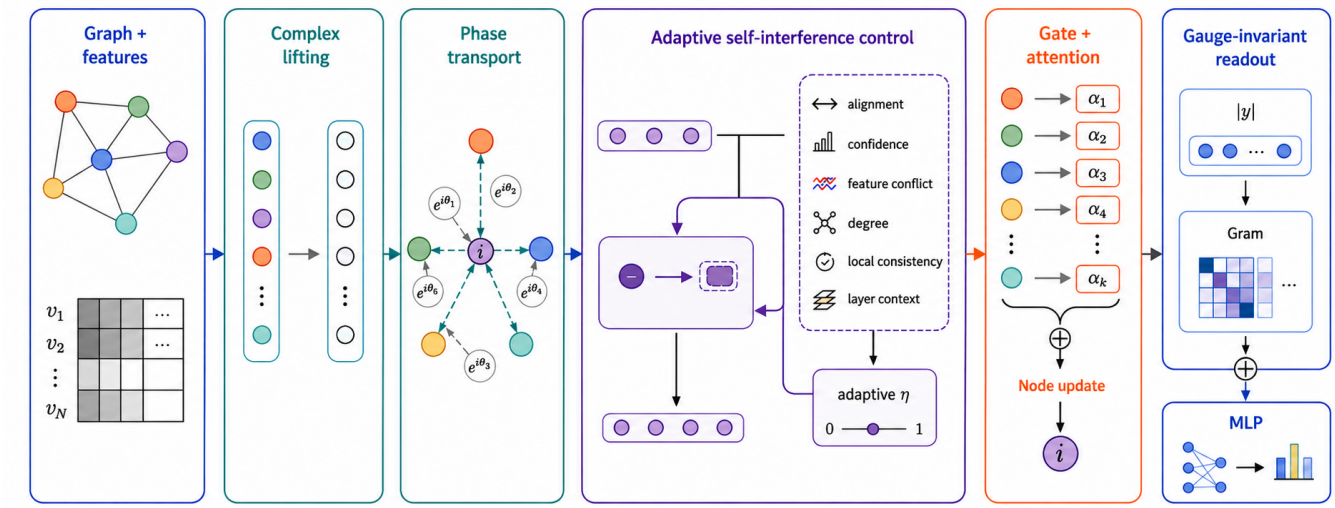

Figure 1: Overview of ASCII. Real node features are lifted into complex states. Each directed route transports a source message into the target frame, computes label-safe reliability descriptors, predicts a bounded cancellation coefficient, removes only the target-parallel component, and then applies gating and attention. The final classifier uses magnitude and Gram features so that logits do not depend on arbitrary node-wise complex frames.

This operator extracts the component of a transported message that is parallel to the current target state up to the regularization factor. The target-parallel component directly reinforces the belief already represented by $h_i^{(\ell)}$, while the orthogonal component can still contain new information about a heterophilic neighbor. ASCII replaces a fixed cancellation strength with a bounded local coefficient:

$$\eta_{ji}^{(m,\ell)} = \eta_{max}\sigma\left((w_m^{(\ell)})^\top z_{ji}^{(m,\ell)} + b_m^{(\ell)}\right), \tag{6}$$

where $0 \le \eta_{max} \le 1$. This leads to the following equation:

$$r_{j\to i}^{(m,\ell)} = \widetilde{h}_{j\to i}^{(m,\ell)} - \eta_{ji}^{(m,\ell)}\Pi_\epsilon(h_i^{(\ell)})\widetilde{h}_{j\to i}^{(m,\ell)}, \tag{7}$$

The descriptor is given by

$$z_{ji}^{(m,\ell)} = [\rho_{ji}^{(m,\ell)}, \chi_{ji}^{(m,\ell)}, \delta_{ji}^x, \delta_{ji}^d, \widehat{q}_i, \widehat{q}_j, \ell/L]. \tag{8}$$

Each entry is real-valued and invariant to the local complex frame. The descriptor groups complementary evidence: transported complex alignment, raw feature conflict, degree-role asymmetry, local inconsistency, and layer context. This makes the adaptive coefficient interpretable as a route-level reliability estimate, not as a hidden label proxy. A small coefficient keeps compatible reinforcement, while a large coefficient removes target-parallel evidence from routes that appear unreliable under these label-safe signals. The alignment and confidence terms are

$$s_{ji}^{(m,\ell)} = (Q_m^{(\ell)}h_i^{(\ell)})^H\widetilde{h}_{j\to i}^{(m,\ell)}, \tag{9}$$

$$\rho_{ji}^{(m,\ell)} = \mathrm{Re}\,\frac{s_{ji}^{(m,\ell)}}{\nu_{ji}^{(m,\ell)}}, \qquad \chi_{ji}^{(m,\ell)} = \frac{|s_{ji}^{(m,\ell)}|}{\nu_{ji}^{(m,\ell)}}, \tag{10}$$

where $\nu_{ji}^{(m,\ell)} = \|Q_m^{(\ell)}h_i^{(\ell)}\|_2\|\widetilde{h}_{j\to i}^{(m,\ell)}\|_2 + \epsilon$. Here ρ records signed transported compatibility, so negative values indicate phase disagreement in the target frame. The term χ records the strength of the phase comparison regardless of sign, so a high value means that the alignment or disagreement is geometrically confident. Feature and degree conflicts are

$$\delta_{ji}^x = 1 - \frac{\langle \bar{x}_i, \bar{x}_j\rangle}{\|\bar{x}_i\|_2\|\bar{x}_j\|_2 + \epsilon}, \qquad \delta_{ji}^d = \log\frac{1+d_i}{1+d_j}. \tag{11}$$

Here $\bar{x}_i$ is a fixed normalized feature view, such as original features, frozen embeddings, or their concatenation. The feature conflict term gives a label-free signal for possible semantic mismatch. The degree term helps distinguish hubs, leaves, and role-transition edges, and its sign preserves the direction of the route. The local inconsistency proxy is computed from frozen unary beliefs u_i or unsupervised feature projections:

$$\widehat{q}_i = 1 - \frac{1}{\max\{|N(i)|, 1\}}\sum_{k\in N(i)}\frac{\langle u_i, u_k\rangle}{\|u_i\|_2\|u_k\|_2 + \epsilon}. \tag{12}$$

If unary beliefs are used, they are produced by a feature-only model trained on V_{tr} and then frozen. The frozen model may produce predictions for all nodes, but its training objective sees only training labels. If this auxiliary model is not used, u_i can be a normalized feature projection trained without labels. In both cases, validation and test labels do not enter the descriptor.

4.3 Gating, Attention, and Readout

After cancellation, ASCII recomputes compatibility using the cancelled message. This prevents the gate and attention score from being based on a component that the adaptive module has just suppressed.

$$\rho_{ji,r}^{(m,\ell)} = \mathrm{Re}\,\frac{(Q_m^{(\ell)}h_i^{(\ell)})^H r_{j\to i}^{(m,\ell)}}{\|Q_m^{(\ell)}h_i^{(\ell)}\|_2\|r_{j\to i}^{(m,\ell)}\|_2 + \epsilon}. \tag{13}$$

A phase-consistent gate is

$$\xi_{ji}^{(m,\ell)} = \sigma(c_m^{(\ell)}\rho_{ji,r}^{(m,\ell)} + d_m^{(\ell)}). \tag{14}$$

A residual mixing gate gives a second safeguard against over-cancellation. It lets the model retain the uncancelled message when the local descriptor suggests that cancellation is not useful. This is important on mixed-homophily graphs because reliable and unreliable routes can appear around the same node:

$$g_{ji}^{(m,\ell)} = \sigma((a_m^{(\ell)})^\top z_{ji}^{(m,\ell)} + a_{0,m}^{(\ell)}), \tag{15}$$

$$\widehat{m}_{j\to i}^{(m,\ell)} = g_{ji}^{(m,\ell)} \xi_{ji}^{(m,\ell)} r_{j\to i}^{(m,\ell)} + (1 - g_{ji}^{(m,\ell)}) \widetilde{h}_{j\to i}^{(m,\ell)}. \tag{16}$$

Attention is computed in the target frame. The first term measures the strength of the target-frame interaction, and the second term keeps its signed phase consistency:

$$\ell_{ji}^{(m,\ell)} = \gamma_m^{(\ell)} \left[\lambda \frac{|\widetilde{s}_{ji}^{(m,\ell)}|}{\sqrt{d}} + (1-\lambda)\,\mathrm{Re}\,\frac{\widetilde{s}_{ji}^{(m,\ell)}}{\|Q_m^{(\ell)} h_i^{(\ell)}\|_2 \|\widehat{m}_{j\to i}^{(m,\ell)}\|_2 + \epsilon} \right], \tag{17}$$

where $\widetilde{s}_{ji}^{(m,\ell)} = (Q_m^{(\ell)} h_i^{(\ell)})^H \widehat{m}_{j\to i}^{(m,\ell)}$. With $\alpha_{ji}^{(m,\ell)} = \mathrm{softmax}_j(\ell_{ji}^{(m,\ell)})$, the update is

$$h_i^{(\ell+1)} = \mathrm{modReLU}\left(\mathrm{NodeNorm}\left(h_i^{(\ell)} + \sum_m \sum_{j\in N(i)} \alpha_{ji}^{(m,\ell)} \widehat{m}_{j\to i}^{(m,\ell)} \right) \right). \tag{18}$$

NodeNorm uses real magnitude statistics and rescales complex states by real factors. Therefore, normalization does not choose a complex frame and does not break equivariance. The final readout uses a learnable projection $R \in \mathbb{C}^{r\times d}$ and $p_i = R h_i^{(L)}$:

$$\phi_i = [|p_i| \| \mathrm{vec}(\mathrm{Re}(p_i p_i^H)) \| \mathrm{vec}(\mathrm{Im}(p_i p_i^H)) \| \|h_i^{(L)}\|_2]. \tag{19}$$

The logits are $o_i = \mathrm{MLP}_{cls}(\mathrm{LayerNorm}(\phi_i))$. Magnitudes retain channel energy, and the Gram matrix retains projected relative phase relations. Raw real and imaginary coordinates are not used by the classifier, so the readout is aligned with the gauge symmetry of propagation. This matters because two equivalent complex frames should produce the same prediction. The rank r controls the $O(r^2)$ readout size.

4.4 Objective Function

The training objective is defined as

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda_{JS}\mathcal{L}_{JS} + \lambda_\eta \mathcal{L}_\eta. \tag{20}$$

$\mathcal{L}_{CE}$ is evaluated only on V_{tr}. The Jensen-Shannon consistency term uses two stochastic forward passes and is optional. It regularizes predictive stability from model outputs and does not add label information outside V_{tr}. The adaptive-coefficient stabilizer is

$$\mathcal{L}_\eta = \frac{1}{M|E|L} \sum_{\ell,m,e} (\eta_e^{(m,\ell)} - \bar{\eta})^2. \tag{21}$$

Setting $\lambda_\eta = 0$ gives a purely data-driven adaptive coefficient. A nonzero value discourages the coefficient from collapsing to a constant extreme during early training. The adaptive module preserves sparse message passing with $O(MEd)$ work per layer after head projection. Compared with fixed-strength SIC, the additional route-level computation is a scalar descriptor score and a scalar gate. The readout costs $O(nrd + nr^2)$, where r is small in practice.

5 Theoretical Analysis

The following results clarify why adaptive cancellation can be added to a complex propagation stack without breaking gauge symmetry or amplifying the self-parallel component. They assume complex-linear projections, $0 \le \eta_{ji}^{(m,\ell)} \le \eta_{max} \le 1$, unit-modulus transports, and gates in $[0,1]$. Each statement is followed by its proof.

Proposition 1 (Descriptor invariance). *Under local $U(1)$ transformations, $z_{ji}^{(m,\ell)}$ and $\eta_{ji}^{(m,\ell)}$ are unchanged.*

Proof. Complex-linear projections commute with scalar phase multiplication. Under the gauge transformation, both $Q_m^{(\ell)} h_i^{(\ell)}$ and $\widetilde{h}_{j\to i}^{(m,\ell)}$ are multiplied by $e^{\iota\psi_i}$ in the target frame. Their Hermitian product, normalized real part, absolute value, and norms are unchanged. Feature conflict, degree descriptors, unary conflict proxies, and layer index are real quantities outside the local complex frame. The descriptor and the real sigmoid coefficient are invariant.

Proposition 2 (Adaptive directional attenuation). *Let $\widetilde{h} = \widetilde{h}_\parallel + \widetilde{h}_\perp$, where $\widetilde{h}_\parallel \in \mathrm{span}\{h_i\}$ and $\langle h_i, \widetilde{h}_\perp\rangle_{\mathbb{C}} = 0$. Then*

$$r = \left(1 - \eta_{ji} \frac{\|h_i\|_2^2}{\|h_i\|_2^2 + \epsilon}\right) \widetilde{h}_\parallel + \widetilde{h}_\perp. \tag{22}$$

Thus cancellation attenuates only the self-parallel component.

Proof. Since $\widetilde{h}_\parallel \in \mathrm{span}\{h_i\}$,

$$\Pi_\epsilon(h_i)\widetilde{h}_\parallel = \frac{\|h_i\|_2^2}{\|h_i\|_2^2 + \epsilon} \widetilde{h}_\parallel. \tag{23}$$

The orthogonality condition gives $\Pi_\epsilon(h_i)\widetilde{h}_\perp = 0$. Applying $I - \eta_{ji}\Pi_\epsilon(h_i)$ gives the stated decomposition.

Proposition 3 (Self-parallel energy non-amplification). *Let P_i be the orthogonal projector onto $\mathrm{span}\{h_i\}$ when $h_i \neq 0$, and let $P_i = 0$ when $h_i = 0$. For $E_\parallel(x) = \|P_i x\|_2^2$ and $\eta_{ji} \in [0,1]$,*

$$E_\parallel((I - \eta_{ji}\Pi_\epsilon(h_i))\widetilde{h}) \le E_\parallel(\widetilde{h}). \tag{24}$$

Proof. If $h_i = 0$, then $P_i = 0$ and the claim is immediate. Otherwise,

$$\Pi_\epsilon(h_i) = \lambda_i P_i, \qquad \lambda_i = \frac{\|h_i\|_2^2}{\|h_i\|_2^2 + \epsilon} \in [0,1). \tag{25}$$

Since $P_i \Pi_\epsilon(h_i) = \lambda_i P_i$,

$$P_i(I - \eta_{ji}\Pi_\epsilon(h_i))\widetilde{h} = (1 - \eta_{ji}\lambda_i) P_i \widetilde{h}. \tag{26}$$

Because $0 \le \eta_{ji} \le 1$ and $0 \le \lambda_i < 1$, we have $|1 - \eta_{ji}\lambda_i| \le 1$. Therefore,

$$\|P_i(I - \eta_{ji}\Pi_\epsilon(h_i))\widetilde{h}\|_2^2 \le \|P_i\widetilde{h}\|_2^2. \tag{27}$$

Proposition 4 (Bounded aggregation). *For every node i and head m,*

$$\left\| \sum_{j\in N(i)} \alpha_{ji}^{(m,\ell)} \widehat{m}_{j\to i}^{(m,\ell)} \right\|_2 \le \|W_m^{(\ell)}\|_2 \max_{j\in N(i)} \|h_j^{(\ell)}\|_2. \tag{28}$$

Proof. Unit-modulus transport preserves norm, giving $\|\widetilde{h}_{j\to i}^{(m,\ell)}\|_2 \le \|W_m^{(\ell)}\|_2 \|h_j^{(\ell)}\|_2$. The eigenvalues of $I - \eta_{ji}\Pi_\epsilon(h_i)$ lie in $[0,1]$, so

Table 1: Node classification results on nine benchmark datasets. Bold indicates the top-3 performance.

Method Homophily	Cora 0.81	Citeseer 0.74	Pubmed 0.80	Actor 0.22	Chameleon 0.23	Squirrel 0.22	Cornell 0.11	Texas 0.06	Wisconsin 0.16
GCN [24]	81.4±0.71	67.5±0.70	79.5±0.47	20.3±0.46	54.9±0.59	31.1±0.71	39.9±0.79	57.0±0.90	49.0±0.78
GAT [38]	82.6±0.55	68.4±0.83	79.9±0.45	22.8±0.41	54.4±0.84	31.0±0.93	42.6±0.80	58.8±1.01	50.2±0.97
H2GCN [45]	80.3±0.52	68.5±0.76	78.8±0.37	25.9±1.07	53.1±0.88	31.2±0.68	55.0±1.15	66.1±1.27	62.0±1.25
SIGN [15]	82.4±0.57	68.6±0.70	79.7±0.43	22.6±0.40	54.2±0.80	31.2±0.94	42.4±0.85	58.8±0.95	50.4±0.88
GCNII [6]	82.2±0.64	67.8±1.21	79.4±0.52	26.2±1.22	54.0±0.77	30.8±0.91	56.0±1.27	69.1±1.34	63.9±1.29
MagNet [42]	83.8±0.56	68.9±0.67	80.1±0.40	26.4±0.97	56.9±1.34	32.4±1.15	55.1±1.31	65.3±1.46	61.7±1.54
GPRGNN [7]	82.0±0.59	70.1±0.91	79.4±0.57	25.2±0.89	55.8±0.81	30.6±0.63	51.4±1.36	60.7±1.28	63.1±1.21
FAGCN [3]	82.8±0.66	70.4±1.20	79.8±0.55	26.8±1.24	54.8±0.81	31.2±0.87	56.8±1.22	69.7±1.41	64.3±1.25
ACM-GCN [30]	82.9±0.70	70.7±0.81	80.0±0.45	25.9±1.02	56.6±1.40	32.1±1.05	55.1±1.35	65.9±1.52	62.1±1.45
GloGNN [28]	82.9±0.45	70.9±0.48	**80.3±0.31**	27.0±0.73	53.9±0.70	31.0±0.82	48.8±1.15	62.5±1.21	60.2±1.12
Auto-HeG [44]	82.4±1.07	70.2±1.36	80.1±0.27	26.5±0.99	54.3±1.33	31.7±1.11	53.9±1.03	67.4±1.65	64.0±1.49
DirGNN [35]	**84.6±0.61**	71.5±0.88	80.2±0.43	27.5±0.95	59.8±1.45	35.2±1.13	**57.9±1.80**	68.8±1.57	63.0±1.33
PCNet [26]	83.4±0.77	70.8±1.15	80.0±0.29	26.6±0.90	57.6±1.65	31.8±0.58	54.1±1.02	62.5±1.16	60.5±1.13
TFE-GNN [14]	**84.1±0.72**	**71.7±1.14**	**80.3±0.30**	28.1±0.81	**60.2±1.61**	**36.0±0.59**	53.7±1.07	63.8±1.11	62.5±1.19
CGNN [46]	83.9±0.70	70.5±1.23	80.1±0.51	26.5±1.17	59.1±0.78	34.4±0.97	57.4±1.25	70.3±1.36	**64.9±1.22**
L2DGCN [13]	82.4±0.99	71.5±0.47	79.8±0.20	**31.3±0.35**	53.1±0.37	35.4±0.52	51.5±3.28	**76.7±2.77**	**65.8±3.01**
GESC [8]	**84.9±0.54**	**72.1±0.51**	**80.4±0.10**	**30.4±0.66**	**65.0±1.29**	**37.9±0.36**	**59.4±1.93**	**74.5±1.52**	**66.6±2.14**
ASCII (ours)	83.8±0.40	**72.4±0.76**	**80.6±0.34**	**30.9±1.58**	**66.5±2.10**	**38.6±0.44**	**62.2±1.69**	**75.7±1.81**	64.7±3.53

$\|r_{j\to i}^{(m,\ell)}\|_2 \le \|\widetilde{h}_{j\to i}^{(m,\ell)}\|_2$. Since $g_{ji}^{(m,\ell)}$ and $\xi_{ji}^{(m,\ell)}$ are in $[0,1]$, the mixed message satisfies

$$\|\widehat{m}_{j\to i}^{(m,\ell)}\|_2 \le g_{ji}^{(m,\ell)}\xi_{ji}^{(m,\ell)}\|r_{j\to i}^{(m,\ell)}\|_2+(1-g_{ji}^{(m,\ell)})\|\widetilde{h}_{j\to i}^{(m,\ell)}\|_2 \le \|\widetilde{h}_{j\to i}^{(m,\ell)}\|_2. \tag{29}$$

The attention weights are nonnegative and sum to one over $N(i)$, giving the result. The statement is per head. Summing the same bound over heads gives

$$\left\|\sum_m \sum_{j\in N(i)} \alpha_{ji}^{(m,\ell)} \widehat{m}_{j\to i}^{(m,\ell)}\right\|_2 \le \sum_m \|W_m^{(\ell)}\|_2 \max_{j\in N(i)} \|h_j^{(\ell)}\|_2. \tag{30}$$

Theorem 5 (Gauge-equivariant propagation and invariant readout)**.** *Assume the gauge action in Section 3, complex-linear projections without additive complex bias, unit-modulus transports,* $0 \le \eta_{ji}^{(m,\ell)} \le 1$*, and phase-preserving NodeNorm and modReLU. Then each ASCII propagation layer is gauge-equivariant. The readout feature in Eq. 19 is gauge-invariant, and the logits are unchanged by local* $U(1)$ *rephasing.*

Proof. Under the gauge action,

$$\widetilde{h}_{j\to i}^{(m,\ell)} = U_{ji}^{(m,\ell)} W_m^{(\ell)} h_j^{(\ell)} \mapsto e^{\imath(\psi_i-\psi_j)} U_{ji}^{(m,\ell)} W_m^{(\ell)} e^{\imath\psi_j} h_j^{(\ell)} = e^{\imath\psi_i}\widetilde{h}_{j\to i}^{(m,\ell)}. \tag{31}$$

The self-alignment operator is unchanged because

$$\Pi_\epsilon(e^{\imath\psi_i} h_i^{(\ell)}) = \frac{e^{\imath\psi_i} h_i^{(\ell)} (e^{\imath\psi_i} h_i^{(\ell)})^H}{\|e^{\imath\psi_i} h_i^{(\ell)}\|_2^2 + \epsilon} = \Pi_\epsilon(h_i^{(\ell)}). \tag{32}$$

By Proposition 1, $z_{ji}^{(m,\ell)}$ and $\eta_{ji}^{(m,\ell)}$ are invariant. Therefore,

$$r_{j\to i}^{(m,\ell)} \mapsto e^{\imath\psi_i} r_{j\to i}^{(m,\ell)}. \tag{33}$$

The recomputed compatibility $\rho_{ji,r}^{(m,\ell)}$, the phase-consistent gate $\xi_{ji}^{(m,\ell)}$, the residual gate $g_{ji}^{(m,\ell)}$, and the attention logit $\ell_{ji}^{(m,\ell)}$ are invariant because they are formed from Hermitian products, real parts, magnitudes, norms, and invariant descriptors after both complex arguments have received the same target-frame phase. Hence the softmax weight $\alpha_{ji}^{(m,\ell)}$ is invariant. It follows that

$$\widehat{m}_{j\to i}^{(m,\ell)} \mapsto e^{\imath\psi_i} \widehat{m}_{j\to i}^{(m,\ell)} \tag{34}$$

and the aggregated message at node i transforms by the same factor $e^{\imath\psi_i}$. The residual term $h_i^{(\ell)}$ also transforms by $e^{\imath\psi_i}$, so the pre-activation state is gauge-equivariant. NodeNorm and modReLU preserve this transformation because they use real magnitude statistics and scalar phase-preserving rescaling. For the readout, $p_i = Rh_i^{(L)}$ transforms as $p_i \mapsto e^{\imath\psi_i} p_i$. Therefore, $|p_i|$, $p_i p_i^H$, and $\|h_i^{(L)}\|_2$ are unchanged. Thus ϕ_i is gauge-invariant. Since LayerNorm and MLP_{cls} receive the same real feature vector ϕ_i, the logits are unchanged.

6 Experiments

This section is organized around two experimental blocks. The first block reports the main node classification comparison on standard benchmark graphs. The second block reports the large graph analysis on OGB node property prediction datasets. The goal is to evaluate whether route-wise adaptive cancellation gives a stronger structural encoder than fixed-strength complex propagation under the same feature source, split protocol, and model-selection rule. All other ablations and diagnostics are kept outside the main experimental section to keep the workshop version focused.

6.1 Main Results

Table 1 summarizes the primary benchmark comparison. The baselines use identical preprocessing, splits, seed count, validation-based model selection, and test-time reporting across all methods. The validation set selects hyperparameters, and the test set is used only once for the final report. The default ASCII grid includes $\eta_{max} \in \{0.5, 0.75, 1.0\}$, $\lambda \in \{0.25, 0.5, 0.75\}$, $r \in \{8, 16\}$, and

Table 2: Integrated large graph analysis on Open Graph Benchmark (OGB) datasets. We compare the original self-interference cancellation (GESC) [8] and our (ASCII) on a single A100 GPU.

Dataset	Nodes	Edges	Prediction task	Split	Metric	GESC [8]			**ASCII (ours)**		
						Score	Time	Memory	Score	Time	Memory
ogbn-arxiv	169,343	1,166,243	40-class paper topic	Time	Accuracy	74.8	44.3s	192MB	76.2	99.8s	248MB
ogbn-proteins	132,534	39,561,252	112-label protein function	Species	ROC-AUC	77.6	27.4s	1.49GB	81.7	62.4s	1.86GB
ogbn-mag	1,939,743	21,111,007	349-class paper venue	Time	Accuracy	44.1	300s	270MB	46.5	706s	356MB

$\lambda_\eta \in \{0, 10^{-4}, 10^{-3}\}$. The key comparison is ASCII against GESC because both methods use complex transport and self-interference cancellation. GESC applies a fixed cancellation strength, while ASCII predicts route-, head-, and layer-specific coefficients from label-safe descriptors. This gives ASCII two practical advantages. It can keep useful target-parallel reinforcement on compatible routes through the residual path, and it can suppress unreliable target-parallel buildup before attention assigns aggregation weights. The comparison is most informative on heterophilic and mixed-homophily graphs, where local route reliability changes sharply across neighborhoods. Across these settings, the results support the central structural claim: replacing a global cancellation strength with route-wise adaptive control improves the reliability of complex propagation under the same input feature source.

6.2 Large Graph Analysis on OGB Datasets

The large graph block evaluates scalability and official split performance on OGB node property prediction datasets [20]. Table 2 reports dataset scale, prediction task, split type, metric, score, training time, and peak memory on a single NVIDIA A100 GPU. ASCII improves over GESC on all three reported OGB datasets. These results show that adaptive cancellation can improve predictive quality on large homogeneous OGB graphs, with additional runtime cost from route-level descriptor scoring and gating. Overall, the OGB results indicate a clear accuracy or ROC-AUC gain on the two completed scored datasets, a consistent increase in training time of about ×2.25, and mixed memory behavior across datasets. This pattern is consistent with the design of ASCII. The sparse transport backbone is retained, while each route receives an additional bounded scalar cancellation coefficient and gate.

7 Conclusion

This paper presents ASCII, a front-end-agnostic structural encoder for reliable graph learning in the large-model era. Rather than proposing another large-model backbone, ASCII addresses the graph-side bottleneck that remains after semantic node features or fused representations are fixed: unreliable routes can still amplify noise, heterophily, or role mismatch during message passing. ASCII learns bounded local cancellation coefficients from gauge-invariant and label-safe descriptors and applies adaptive cancellation before attention. Together with a magnitude-Gram readout, the model preserves gauge symmetry from propagation to prediction. Theoretical analysis establishes descriptor invariance, self-parallel non-amplification, bounded sparse aggregation, and invariant readout. Experiments on standard benchmarks and OGB datasets evaluate the structural mechanism under matched feature sources, supporting ASCII as a reusable graph encoder for hybrid graph foundation model.

GenAI Usage Disclosure

The authors used AI-based tools only for grammar checking and proofreading. All technical ideas, mathematical derivations, experiments, results, and claims were produced and verified by the authors.

References

[1] Sami Abu-El-Haija, Bryan Perozzi, Amol Kapoor, Nazanin Alipourfard, Kristina Lerman, Hrayr Harutyunyan, Greg Ver Steeg, and Aram Galstyan. 2019. MixHop: Higher-Order Graph Convolutional Architectures via Sparsified Neighborhood Mixing. In *International Conference on Machine Learning (ICML)*.

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

[3] Deyu Bo, Xiao Wang, and Chuan Shi. 2021. Beyond low-frequency information in graph convolutional networks. In *AAAI*.

[4] Johannes Brandstetter, Max Welling, and Maurice Weiler. 2022. Geometric and Physical Quantities Improve E(3) Equivariant Message Passing. In *International Conference on Learning Representations (ICLR)*.

[5] Michael M. Bronstein, Joan Bruna, Taco S. Cohen, and Petar Veličković. 2021. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. *arXiv preprint arXiv:2104.13478* (2021).

[6] Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. 2020. Simple and deep graph convolutional networks. In *International conference on machine learning*. PMLR, 1725–1735.

[7] Eli Chien, Chih-Kuan Pan, Wei-Cheng Peng, and Olgica Milenkovic. 2021. Adaptive Universal Generalized PageRank Graph Neural Network. In *ICLR*.

[8] Yoonhyuk Choi, Jiho Choi, and Chong-Kwon Kim. 2026. Gauge-Equivariant Graph Networks via Self-Interference Cancellation. *ICML* (2026).

[9] Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. 2019. Gauge Equivariant Convolutional Networks and the Icosahedral CNN. In *ICML*.

[10] Pim de Haan, Taco S. Cohen, and Max Welling. 2021. Gauge Equivariant Mesh CNNs: Anisotropic convolutions on geometric graphs. In *International Conference on Learning Representations (ICLR)*.

[11] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *NeurIPS*.

[12] Tyler Derr, Yao Ma, and Jiliang Tang. 2018. Signed graph convolutional networks. In *2018 IEEE international conference on data mining (ICDM)*. IEEE, 929–934.

[13] Shifei Ding, Jian Zhang, Lili Guo, Xuan Li, et al. 2025. L2DGCN: Learnable Enhancement and Label Selection Dynamic Graph Convolutional Networks for Mitigating Degree Bias. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

[14] Rui Duan, Mingjian Guang, Junli Wang, Chungang Yan, Hongda Qi, Wenkang Su, Can Tian, and Haoran Yang. 2024. Unifying homophily and heterophily for spectral graph neural networks via triple filter ensembles. *Advances in Neural Information Processing Systems* 37 (2024), 93540–93567.

[15] Fabrizio Frasca, Emanuele Rossi, Davide Eynard, Ben Chamberlain, Michael Bronstein, and Federico Monti. 2020. Sign: Scalable inception graph neural networks. *arXiv preprint arXiv:2004.11198* (2020).

[16] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. 2017. Neural message passing for quantum chemistry. In *ICML*.

[17] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

[18] Mingguo He, Zhewei Wei, Zengfeng Huang, and Hongteng Xu. 2021. BernNet: Learning Arbitrary Graph Spectral Filters via Bernstein Approximation. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[19] Yixuan He, Michael Perlmutter, Gesine Reinert, and Mihai Cucuringu. 2022. MSGNN: A Spectral Graph Neural Network Based on a Novel Magnetic Signed Laplacian. In *Proceedings of the First Learning on Graphs Conference (PMLR 198)*. 40:1–40:39.

[20] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. 2020. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems* 33 (2020), 22118–22133.

[21] Jing Huang, Jie Shen, Ruochi Hou, Shuchang Liu, and Xiaofei Ma. 2019. Signed graph attention networks. In *ICLR*.

[22] Kexin Huang, Ying Jin, Emmanuel Candes, and Jure Leskovec. 2023. Uncertainty quantification over graph with conformalized graph neural networks. *Advances in Neural Information Processing Systems* 36 (2023), 26699–26721.

[23] Wei Jin, Yao Ma, Xiaorui Liu, Xianfeng Tang, Suhang Wang, and Jiliang Tang. 2020. Graph structure learning for robust graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 66–74.

[24] Thomas N Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. *ICLR* (2017).

[25] Taewook Ko, Yoonhyuk Choi, and Chong-Kwon Kim. 2023. A spectral graph convolution for signed directed graphs via magnetic Laplacian. *Neural Networks* 164 (2023), 562–574.

[26] Bingheng Li, Erlin Pan, and Zhao Kang. 2024. Pc-conv: Unifying homophily and heterophily with two-fold filtering. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 38. 13437–13445.

[27] Qimai Li, Zhichao Han, and Xiao-Ming Wu. 2018. Deeper insights into graph convolutional networks for semi-supervised learning. In *AAAI*.

[28] Xiang Li, Renyu Zhu, Yao Cheng, Caihua Shan, Siqiang Luo, Dongsheng Li, and Weining Qian. 2022. Finding global homophily in graph neural networks when meeting heterophily. In *International conference on machine learning*. PMLR, 13242–13256.

[29] Derek Lim, Felix Hohne, Xiuyu Li, Sijia Linda Huang, Vaishnavi Gupta, Omkar Bhalerao, and Ser Nam Lim. 2021. Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods. *Advances in neural information processing systems* 34 (2021), 20887–20902.

[30] Sitao Luan, Chenqing Hua, Qincheng Lu, Jiaqi Zhu, Mingde Zhao, Shuyuan Zhang, Xiao-Wen Chang, and Doina Precup. 2022. Revisiting heterophily for graph neural networks. *Advances in neural information processing systems* 35 (2022), 1362–1375.

[31] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 27, 1 (2001), 415–444.

[32] Mark EJ Newman. 2002. Assortative mixing in networks. *Physical review letters* 89, 20 (2002), 208701.

[33] Kenta Oono and Taiji Suzuki. 2019. Graph neural networks exponentially lose expressive power for node classification. In *ICLR*.

[34] Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. 2020. Geom-GCN: Geometric Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.

[35] Emanuele Rossi, Bertrand Charpentier, Francesco Di Giovanni, Fabrizio Frasca, Stephan Günnemann, and Michael M Bronstein. 2024. Edge directionality improves learning on heterophilic graphs. In *Learning on graphs conference*. PMLR, 25–1.

[36] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In *ICLR*.

[37] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *ICML*.

[38] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *ICLR*.

[39] Junfu Wang, Yuanfang Guo, Liang Yang, and Yunhong Wang. 2024. Understanding Heterophily for Graph Neural Networks. In *Proceedings of the 41st International Conference on Machine Learning (PMLR 235)*. 50489–50529.

[40] Xiyuan Wang and Muhan Zhang. 2022. How powerful are spectral graph neural networks. In *International conference on machine learning*. PMLR, 23341–23362.

[41] Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. 2018. Representation learning on graphs with jumping knowledge networks. In *International conference on machine learning*. pmlr, 5453–5462.

[42] Xiaorong Zhang, Jiajin Bu, Zhao Zhu, Qi Xuan, and Zhi Yu. 2021. MagNet: A Neural Network for Directed Graphs. In *NeurIPS*.

[43] Xiang Zhang and Marinka Zitnik. 2020. Gnnguard: Defending graph neural networks against adversarial attacks. *Advances in neural information processing systems* 33 (2020), 9263–9275.

[44] Xin Zheng, Miao Zhang, Chunyang Chen, Qin Zhang, Chuan Zhou, and Shirui Pan. 2023. Auto-heg: Automated graph neural network on heterophilic graphs. In *Proceedings of the ACM Web Conference 2023*. 611–620.

[45] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. 2020. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in neural information processing systems* 33 (2020), 7793–7804.

[46] Wei Zhuo, Han Yu, Guang Tan, and Xiaoxiao Li. 2025. Commute Graph Neural Networks. In *Proceedings of the 42nd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 267)*, Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (Eds.). PMLR, 80612–80628. https://proceedings.mlr.press/v267/zhuo25a.html