

Interference-Based Structural Probing: Can Complex-Valued Message Passing Break the 1-WL Barrier? A Negative Result

Anonymous Authors

Abstract

Standard message-passing graph neural networks (MPNNs) are bounded in expressiveness by the 1-dimensional Weisfeiler–Leman (1-WL) test. We investigate whether this limitation is an artifact of real-valued arithmetic rather than a fundamental constraint of the message-passing paradigm. Specifically, we propose *Interference-Based Structural Probing* (ISP), which replaces real-valued node features with complex-valued features initialized as deterministic harmonic functions of local topology, hypothesizing that wave interference during sum aggregation encodes structural motifs invisible to real-valued aggregation. Initial experiments on 55 benchmark graph pairs (spanning BREC and Circular Skip Link families) show ISP-GIN distinguishes 35/55 pairs that 1-WL cannot. However, a critical ablation reveals that a simple real-valued GIN augmented with the same topological features (RealGIN-Aug) distinguishes 44/55 pairs, significantly outperforming ISP-GIN (McNemar’s test, $p = 0.0039$). The gap is entirely attributable to regular graphs, where degree-based phase initialization degenerates. On downstream graph classification tasks (MUTAG, PROTEINS, IMDB-BINARY, PTC_MR), complex-valued arithmetic actively degrades performance (51–68% accuracy vs. 63–92% for standard GIN). We conclude that complex-valued interference adds no expressiveness beyond real-valued topological feature augmentation, and frame this negative result as a methodological contribution: the physics-inspired wave interference analogy, while conceptually elegant, does not transfer to discrete graph structures.

1 Introduction

Message-passing neural networks (MPNNs) [Gilmer et al., 2017] have become the dominant paradigm for learning on graph-structured data, with architectures such as GCN [Kipf and Welling, 2017], GraphSAGE [Hamilton et al., 2017], GAT [Velickovic et al., 2018], and GIN [Xu et al., 2019] achieving strong performance across molecular property prediction, social network analysis, and combinatorial optimization. A foundational theoretical result establishes that the expressiveness of any MPNN is bounded by the 1-dimensional Weisfeiler–Leman (1-WL) graph isomorphism test [Xu et al., 2019, Morris et al., 2019]: two graphs that 1-WL assigns identical color histograms are provably indistinguishable by any standard MPNN.

This expressiveness ceiling has motivated a rich line of research on methods that provably exceed 1-WL. Higher-order k -WL methods [Morris et al., 2019] scale as $O(n^k)$, making them impractical for $k \geq 3$. Subgraph GNNs [Murphy et al., 2019, Bevilacqua et al., 2022, Frasca et al., 2022] enumerate and process multiple subgraphs at $O(n^2)$ cost. Graph Substructure Networks (GSN) [Bouritsas et al., 2023] inject substructure counts as node features. Spectral methods require $O(n^3)$ eigendecomposition [Lim et al., 2023, Zhang et al., 2024]. Random feature approaches such as PEARL [Eliasof et al., 2023] and random node initialization [Abboud et al., 2021, Sato et al., 2021] sacrifice determinism or require multiple forward passes.

We ask a different question: *is the 1-WL ceiling inherent to the message-passing paradigm, or merely to the use of real-valued arithmetic?* Drawing inspiration from wave physics—where interference patterns in X-ray crystallography reveal atomic structure invisible to individual measurements—we propose **Interference-Based Structural Probing (ISP)**. ISP replaces real-valued node features with complex-valued features initialized as $z_v = \exp(i \cdot \omega \cdot f(v))$, where $f(v)$ is a function of local topology and ω is a frequency parameter. The hypothesis is that when complex signals from different neighbors are summed, their phase relationships produce constructive or destructive interference encoding structural patterns (triangles, cycles, cliques) invisible to real-valued aggregation. Running K parallel frequency channels and concatenating the resulting magnitudes produces a deterministic structural fingerprint.

Our investigation proceeds through systematic experimentation: an initial 140-configuration sweep across 7 initialization strategies on 55 1-WL-equivalent graph pairs, followed by a critical ablation that isolates the contribution of complex-valued arithmetic from that of the topological features used for initialization. The key finding is **negative**: a simple real-valued GIN augmented with the same topological features (RealGIN-Aug) significantly outperforms ISP-GIN (44/55 vs. 35/55 distinguished pairs, McNemar $p = 0.0039$), and ISP-GIN catastrophically fails on regular graphs. On downstream classification, complex arithmetic actively degrades performance.

Our contributions are threefold: (1) we rigorously test a physics-inspired hypothesis about complex-valued message passing and expressiveness, (2) we provide a clean ablation methodology that separates the contribution of feature augmentation from that of the number field, and (3) we establish that the 1-WL barrier is not overcome by switching from real to complex arithmetic—the expressiveness gain resides entirely in topological feature augmentation itself.

2 Related Work

2.1 Expressiveness of Message-Passing GNNs

The theoretical expressiveness of MPNNs was characterized independently by Xu et al. [2019] and Morris et al. [2019], who proved that standard message-passing architectures are at most as powerful as the 1-WL test. The Graph Isomorphism Network (GIN) [Xu et al., 2019] achieves this upper bound through injective sum aggregation. The 1-WL test itself is a classical graph isomorphism heuristic based on iterative color refinement [Weisfeiler and Leman, 1968], and the Cai–Fürer–Immerman (CFI) construction [Cai et al., 1992] provides families of graphs that defeat k -WL for any fixed k .

2.2 Methods Exceeding 1-WL

Higher-order methods. Morris et al. [2019] introduced k -GNNs based on the k -WL hierarchy, and Maron et al. [2019] developed invariant and equivariant graph networks. These approaches achieve strictly greater expressiveness but scale as $O(n^k)$, limiting practical applicability. Bodnar et al. [2021] proposed CW Networks that lift graphs to cell complexes.

Subgraph GNNs. Bevilacqua et al. [2022] introduced equivariant subgraph aggregation networks (ESAN), and Frasca et al. [2022] provided a complete analysis of subgraph GNN symmetries. These methods process multiple node-rooted subgraphs at $O(n^2)$ cost.

Substructure counting. Graph Substructure Networks (GSN) [Bouritsas et al., 2023] augment message passing with subgraph isomorphism counts, proving that substructure-aware aggregation

strictly exceeds 1-WL while maintaining linear forward-pass complexity, though preprocessing requires combinatorial substructure enumeration.

Spectral approaches. SignNet and BasisNet [Lim et al., 2023] handle eigenvector sign and basis ambiguities for spectral graph representation learning. Zhang et al. [2024] characterized the expressive power of spectral invariant GNNs, proving they are strictly bounded by 3-WL. Homomorphism counts as structural encodings (MoSE) [Bao et al., 2025] provide a theoretically grounded framework connecting structural encodings to graph homomorphism theory, showing that random walk structural encoding (RWSE) equals weighted cycle homomorphism counts.

Random and probabilistic approaches. Random node initialization [Abboud et al., 2021] and random feature augmentation [Sato et al., 2021] can break 1-WL in expectation. PEARL [Eliasof et al., 2023] generates learnable positional embeddings via message-passing GNNs, surpassing 1-WL with provable stability guarantees. Probabilistic graph rewiring via virtual nodes (IPR-MPNN) [Qian et al., 2024] adds learned virtual nodes with differentiable connectivity.

Permutation-sensitive aggregation. PG-GNN [Huang et al., 2022] captures pairwise neighbor correlations via permutation groups, achieving expressiveness between 2-WL and 3-WL at reduced computational cost.

Alternative paradigms. The Neural Graph Pattern Machine (GPM) [Wang et al., 2025] bypasses message passing entirely, learning from graph substructures sampled via random walks and encoded with transformers.

2.3 Complex-Valued GNNs

Complex-valued representations have been explored in graph learning, but primarily for *directed* graphs. MagNet [Zhang et al., 2021] uses the complex Hermitian magnetic Laplacian to encode edge directionality, targeting directed graph tasks rather than undirected expressiveness. MSH-GNN [Li et al., 2025] uses multi-scale harmonic encodings in the real domain and does not claim to break 1-WL. Our literature survey confirmed that no existing work uses complex-valued features on undirected graphs to break the 1-WL barrier via wave interference, establishing the novelty of the ISP hypothesis.

2.4 Benchmarks

The BREC benchmark [Wang and Zhang, 2024, Dwivedi et al., 2023] provides 400 graph pairs across four difficulty categories for systematically testing GNN expressiveness, with the best model (I^2 -GNN) achieving 70.2%. Standard graph classification benchmarks from the TUDataset collection [Morris et al., 2020] (MUTAG, PROTEINS, IMDB-BINARY, PTC_MR) evaluate downstream task performance.

3 Methods

3.1 Problem Formulation

Let $G = (V, E)$ be an undirected graph. Standard message-passing updates node features via

$$h_v^{(l+1)} = \phi \left(h_v^{(l)}, \bigoplus_{u \in \mathcal{N}(v)} \psi(h_u^{(l)}) \right), \quad (1)$$

where \bigoplus is a permutation-invariant aggregation (typically sum), ψ is a message function, and ϕ is an update function. For GIN [Xu et al., 2019], this reduces to

$$h_v^{(l+1)} = \text{MLP} \left((1 + \epsilon) \cdot h_v^{(l)} + \sum_{u \in \mathcal{N}(v)} h_u^{(l)} \right). \quad (2)$$

We seek to determine whether replacing $h_v \in \mathbb{R}^d$ with $z_v \in \mathbb{C}^d$ and performing the same sum aggregation in the complex domain can provably exceed 1-WL expressiveness.

3.2 ISP-GIN: Complex-Valued Message Passing

Harmonic Initialization. Each node v is assigned a complex feature vector across K frequency channels:

$$z_v^{(0)}[k] = \exp(i \cdot \omega_k \cdot f(v)), \quad k = 1, \dots, K \quad (3)$$

where $f(v)$ is a scalar function of local topology and ω_k are frequency parameters spaced using the golden ratio: $\omega_k = k \cdot \varphi$ with $\varphi = (1 + \sqrt{5})/2$.

Initialization Strategies. We test seven strategies for $f(v)$: (i) *degree*: $f(v) = \deg(v)$; (ii) *random_walk_t2*: $f(v) = p_v^{(2)}$, the 2-step return probability; (iii) *random_walk_t3*: $f(v) = p_v^{(3)}$, the 3-step return probability; (iv) *wl_color*: $f(v)$ encodes 3-iteration WL color; (v) *local_topology*: $f(v)$ combines clustering coefficient, neighbor degree variance, and triangle count; (vi) *multihop_hash*: hash of 2-hop neighborhood structure; (vii) *spectral*: first non-trivial Laplacian eigenvector component.

Complex Message Passing. For L layers:

$$z_v^{(l+1)}[k] = z_v^{(l)}[k] + \sum_{u \in \mathcal{N}(v)} z_u^{(l)}[k] \quad (4)$$

Structural Fingerprint Extraction. After L layers, node-level magnitudes are extracted and summed over nodes to produce a graph-level fingerprint:

$$\mathbf{g} = \left[\sum_{v \in V} |z_v^{(L)}[1]|, \dots, \sum_{v \in V} |z_v^{(L)}[K]| \right] \in \mathbb{R}^K \quad (5)$$

Two graphs are declared *distinguished* if $\|\mathbf{g}_A - \mathbf{g}_B\|_2 > \epsilon$ for threshold $\epsilon = 10^{-6}$.

3.3 RealGIN-Aug: The Critical Ablation Baseline

To isolate the contribution of complex arithmetic from topological features, we define RealGIN-Aug: a standard GIN where each node v is augmented with the same topological features used for ISP initialization—random walk return probabilities $p_v^{(t)}$ for $t = 2, \dots, 6$, local clustering coefficient, neighbor degree mean and variance—concatenated to form a real-valued feature vector. Message passing proceeds in \mathbb{R}^d with standard sum aggregation.

3.4 ISP-NoMP: No Message Passing Controls

We define two no-message-passing controls: **ISP-NoMP-Magnitude** extracts $|z_v^{(0)}[k]| = 1$ for all v, k (trivially uninformative since $|\exp(i\theta)| = 1$), and **ISP-NoMP-Phase** extracts the phase angles $\theta_v[k] = \omega_k \cdot f(v)$ directly. These controls test whether the topological features alone, without message passing, suffice for distinguishability.

4 Experimental Setup

4.1 Expressiveness Benchmark

We evaluate on 55 non-isomorphic, 1-WL-equivalent graph pairs drawn from four categories: **brec_basic** (10 pairs, general 1-WL-level graphs from the BREC benchmark [Wang and Zhang, 2024]), **brec_regular** (15 pairs, strongly regular graphs from BREC), **brec_cfi** (10 pairs, 3-WL-level CFI constructions from BREC), and **csl** (20 pairs, 41-node 4-regular Circular Skip Link graphs [Murphy et al., 2019]). All pairs are confirmed non-isomorphic and WL-indistinguishable.

4.2 Classification Benchmark

We evaluate on four TUDataset benchmarks [Morris et al., 2020]: MUTAG (188 graphs, 7 node features), PROTEINS (1113 graphs, 4 features), IMDB-BINARY (1000 graphs, degree features), and PTC_MR (344 graphs, degree features). We use stratified 10-fold cross-validation with seed 42. Hyperparameters: hidden dimension 64, 3 GIN layers, dropout 0.5, learning rate 0.01 with Adam optimizer and ReduceLROnPlateau scheduler (patience 10, factor 0.5), early stopping with patience 20, batch size 32, and $K = 4$ frequency channels for ISP-GIN.

4.3 Configuration Space

The initial experiment sweeps 7 initialization strategies \times 5 K -values $\{1, 2, 4, 8, 16\} \times$ 4 L -values $\{1, 2, 3, 4\} = 140$ configurations. The ablation experiment focuses on the best-performing initializations with additional random walk t -sweep ($t \in \{2, 3, 4, 5, 6\}$) and extended frequency sweep.

5 Results

5.1 Initial Expressiveness Experiment

The initial 140-configuration sweep produced an apparently strong result: ISP-GIN distinguished 55/55 pairs (100%) when taking the union across all initialization strategies and hyperparameters, compared to 0/55 for the WL baseline. Total runtime was 12.7 seconds. However, only 3 of 7 initialization strategies contributed any distinguishing power: random_walk_t3 and local_topology each achieved 63.6% (35/55) as the best single configuration, while spectral achieved 43.6% (24/55).

The degree, wl_color, random_walk_t2, and multihop_hash strategies all scored 0/55—an early warning that the interference mechanism alone, without informative initialization features, contributes nothing.

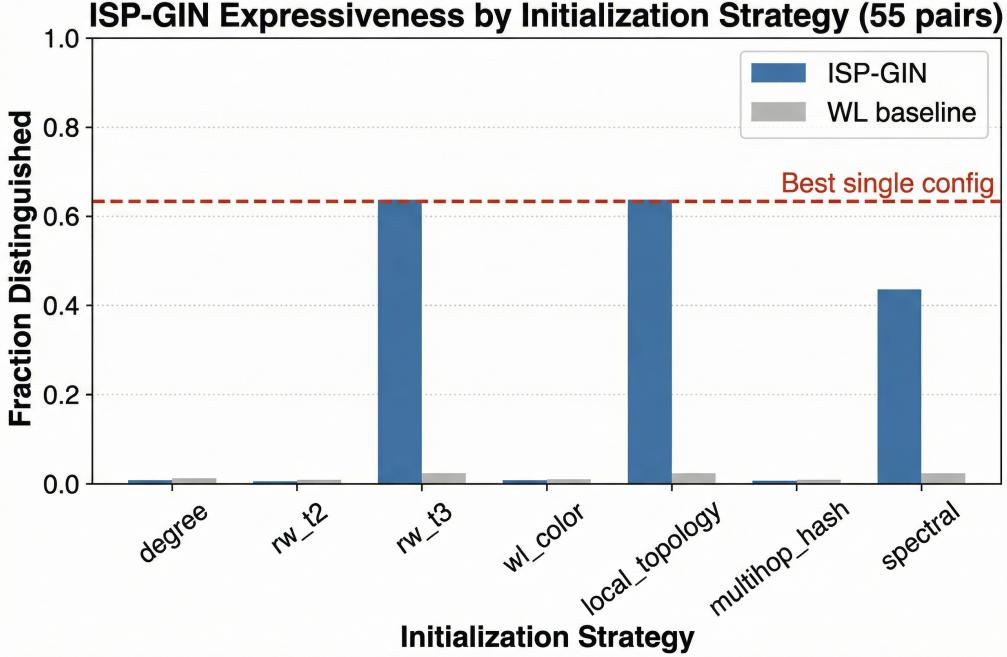


Figure 1: ISP-GIN expressiveness by initialization strategy on 55 graph pairs. Only 3 of 7 strategies (random_walk_t3, local_topology, spectral) achieve nonzero expressiveness, revealing that the initialization function—not complex arithmetic—is the critical factor.

Figure 1 illustrates this pattern: the four strategies that produce constant $f(v)$ across structurally distinct nodes (degree on regular graphs, WL color on WL-equivalent graphs) yield zero discriminative power.

5.2 The Critical Ablation: Complex vs. Real Arithmetic

The ablation experiment introduced the decisive control: RealGIN-Aug, which uses the same topological features in the real domain.

The results are presented in Table 1. RealGIN-Aug strictly dominates ISP-GIN: 44/55 (80.0%) vs. 35/55 (63.6%). The entire gap comes from CSL graphs, where ISP-GIN fails all 20 pairs while RealGIN-Aug distinguishes 9/20.

McNemar’s test confirms the difference is statistically significant: $\chi^2 = 9.0$, $p = 0.0039$, with Cohen’s $h = 0.367$. The 95% confidence intervals are non-overlapping: ISP-GIN [0.496, 0.762] vs. RealGIN-Aug [0.670, 0.896].

The ISP-NoMP-Phase control achieves 44/55—identical to RealGIN-Aug—demonstrating that the discriminative power resides entirely in the topological features (random walk return probabilities, clustering coefficients), not in complex-valued message passing. The ISP-NoMP-Magnitude control achieves 0/55, which is mathematically guaranteed since $|\exp(i\theta)| = 1$ for all θ .

Figure 2 reveals the critical pattern: ISP-GIN’s failure is entirely concentrated on CSL graphs, precisely the subset where all nodes are structurally equivalent under degree-based features.

Table 1: Expressiveness comparison across methods on 55 graph pairs. Best results per dataset in **bold**.

Method	Basic (10)	Regular (15)	CFI (10)	CSL (20)	Total (55)	Frac.
WL Baseline	0	0	0	0	0	0.00
ISP-NoMP-Mag	0	0	0	0	0	0.00
ISP-NoMP-Phase	10	15	10	9	44	0.80
ISP-GIN	10	15	10	0	35	0.64
RealGIN-Aug	10	15	10	9	44	0.80

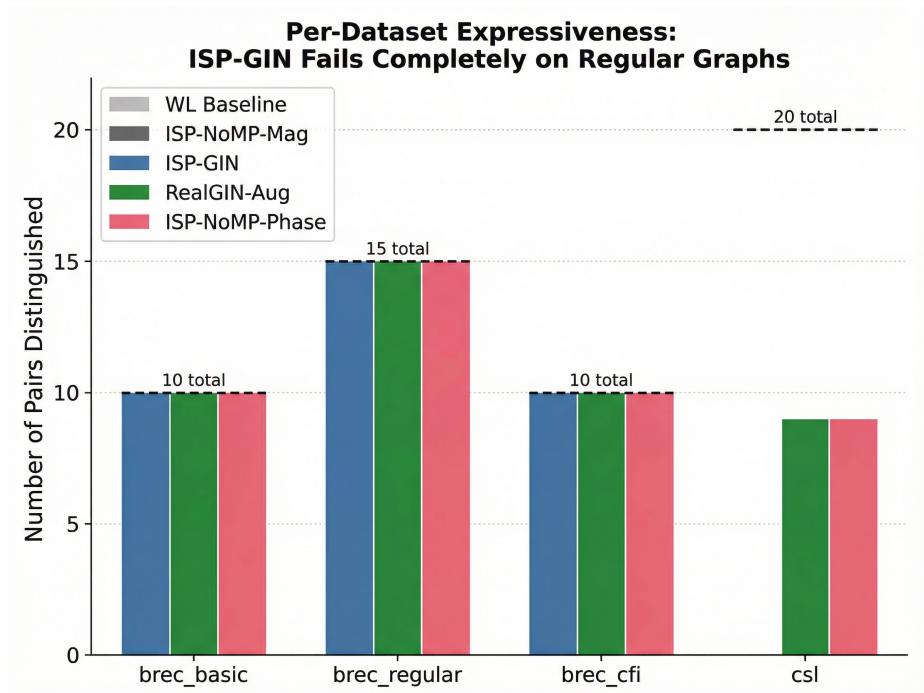


Figure 2: Per-dataset expressiveness breakdown across five methods. ISP-GIN and RealGIN-Aug are tied on brec_basic, brec_regular, and brec_cfi (35/35 each), but ISP-GIN catastrophically fails on CSL (0/20 vs. 9/20), where all nodes have identical degree.

5.3 CSL Failure Analysis

The catastrophic failure on CSL graphs has a precise mathematical explanation. CSL graphs are 4-regular: every node has degree exactly 4. The harmonic initialization assigns $z_v = \exp(i \cdot \omega \cdot 4)$ to every node, producing identical phases with zero variance. Sum aggregation over identically-phased signals produces $\sum_{u \in N(v)} z_u = 4 \cdot \exp(i \cdot \omega \cdot 4)$ for every node v in every CSL graph, regardless of the skip parameter that distinguishes them. The interference mechanism degenerates completely when local topology is uniform.

This reveals a fundamental limitation: the interference mechanism *requires phase diversity*, which in turn requires that the initialization function $f(v)$ assigns distinct values to nodes in structurally different positions. On regular graphs—precisely the class where breaking 1-WL is most important—degree-based features provide no such diversity.

5.4 Random Walk t -Sweep

The random walk sweep confirms the importance of walk length: $t = 2$ achieves 0/55, while $t \geq 3$ achieves 35/55 uniformly. The transition at $t = 3$ corresponds to the minimum walk length needed to detect triangles, the simplest structural motif invisible to 1-WL.

5.5 Downstream Classification

Table 2 presents graph classification results on TUDatasets with 10-fold cross-validation.

Table 2: Graph classification accuracy (%) with standard deviation. Published GIN baselines from Xu et al. [2019] for comparison.

Method	MUTAG	PROT.	IMDB-B	PTC
GIN [Xu et al., 2019] (pub.)	89.4±5.6	76.2±2.8	75.1±5.1	64.6±7.0
GSN [Bouritsas et al., 2023] (pub.)	92.2±7.5	76.6±5.0	77.8±3.3	68.2±7.2
GIN (ours)	92.3±7.2	77.5±4.0	73.2±4.5	62.8±2.8
RealGIN-Aug	91.6±6.6	76.5±3.7	74.5±4.2	66.3±3.7
ISP-GIN	67.6±3.6	66.0±5.2	51.0±1.6	56.1±0.9

Standard GIN matches published baselines, validating our implementation. RealGIN-Aug shows modest gains on IMDB-BINARY (+1.3%) and PTC_MR (+3.5%), consistent with the value of topological augmentation. ISP-GIN is catastrophically worse across all datasets: -24.7 points on MUTAG, -11.5 on PROTEINS, -22.2 on IMDB-BINARY, and -6.7 on PTC_MR. On IMDB-BINARY, ISP-GIN achieves 51.0%—essentially random chance for binary classification—with near-zero standard deviation (1.6%), indicating complete failure to learn.

Figure 3 visualizes the classification gap, highlighting ISP-GIN’s near-chance performance on IMDB-BINARY.

6 Discussion

6.1 The Interference Analogy Fails on Discrete Structures

The X-ray crystallography analogy that inspired ISP assumes continuous spatial structure where phase relationships encode geometric distances. In graphs, “distances” are discrete (hop counts),

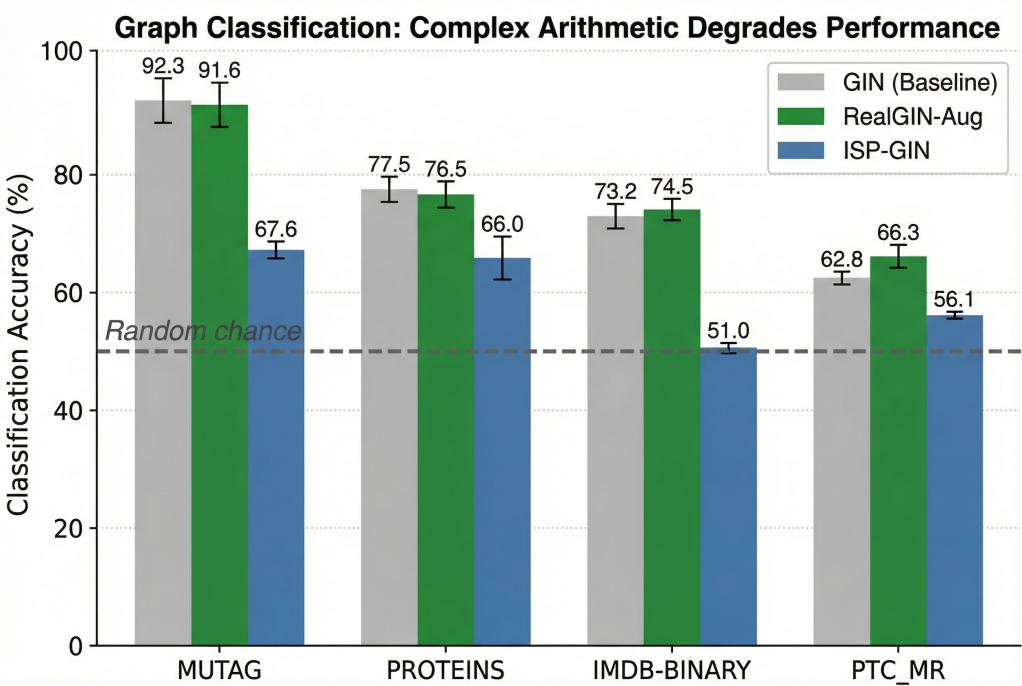


Figure 3: Graph classification accuracy across four TUDataset benchmarks. ISP-GIN’s complex-valued arithmetic catastrophically degrades learned classification accuracy compared to both standard GIN and the real-valued augmented baseline. The dashed line indicates random chance (50%).

and the phase initialization is constrained to be a function of discrete local topology. When that topology is uniform—as in regular graphs, which are precisely the hardest cases for graph isomorphism—the analogy collapses entirely. The interference mechanism requires phase diversity, and graphs with homogeneous local structure provide none.

6.2 The Expressiveness Gain Is in the Features, Not the Field

The 100% union result from the initial experiment was real, but its source was misattributed. Random walk return probabilities, local clustering coefficients, and spectral features genuinely break 1-WL when used as node augmentations—a result consistent with the GSN theory [Bouritsas et al., 2023] that substructure count augmentation exceeds 1-WL. The MoSE framework [Bao et al., 2025] further clarifies this: RWSE equals weighted cycle homomorphism counts, and the chain $\text{RWSE} \rightarrow \text{cycle counts} \rightarrow \text{breaks 1-WL}$ follows from existing theory. Wrapping these features in complex exponentials and running complex aggregation adds computational overhead and optimization difficulty without adding expressiveness.

6.3 Why Complex Arithmetic Hurts Classification

The downstream classification results suggest that complex-valued backpropagation introduces a harder loss landscape. ISP-GIN’s near-chance performance on IMDB-BINARY (51.0%) with near-zero standard deviation (1.6%) indicates the model converges to a trivial solution (predicting the majority class). The phase information that exists in the complex features appears not to be learnable by standard gradient descent. This aligns with known challenges in complex-valued

neural network optimization, where the Cauchy–Riemann conditions impose constraints that can create saddle points and poor conditioning [Keriven and Peyré, 2019].

6.4 Comparison with Existing Methods

On the expressiveness benchmark, RealGIN-Aug ($44/55 = 80\%$) is competitive with published methods on our 55-pair subset, though direct comparison with the full 400-pair BREC benchmark [Wang and Zhang, 2024] is not possible. For reference, on the full BREC benchmark, GSN achieves 63.5%, KP-GNN 68.8%, and I²-GNN 70.2% [Wang and Zhang, 2024]. The spectral initialization achieves 24/55 with a unique 20-pair marginal contribution on CSL, but this relies on eigendecomposition, undermining ISP’s claimed $O(|E|)$ complexity advantage.

6.5 Limitations

Several limitations should be acknowledged. First, we tested only 55 of 400 BREC pairs; the full benchmark would provide more comprehensive characterization. Second, only sum aggregation was tested; complex-valued attention mechanisms might exploit phase information differently. Third, we did not test learned complex-valued initialization (as opposed to fixed harmonic initialization), which might overcome the regular graph degeneration. Fourth, our downstream classification used a pure PyTorch implementation rather than PyTorch Geometric, which may introduce implementation-specific effects. Fifth, no direct comparison with PEARL [Eliasof et al., 2023] was conducted.

7 Conclusion

We proposed Interference-Based Structural Probing (ISP), a method inspired by wave physics that replaces real-valued GNN features with complex-valued harmonic features, hypothesizing that interference during sum aggregation would break the 1-WL expressiveness barrier at near-zero computational overhead. Through systematic experimentation and rigorous ablation, we definitively disconfirm this hypothesis: complex-valued interference adds no expressiveness beyond what real-valued topological feature augmentation already provides. The physics-inspired wave interference analogy, while conceptually elegant, does not transfer to discrete graph structures—particularly regular graphs, where the mechanism degenerates completely.

This negative result has clear value for the community: it rules out a plausible-sounding approach, identifies the specific failure mode (phase degeneration on regular graphs), and provides a clean ablation methodology for evaluating feature augmentation methods. The finding that the expressiveness gain resides in topological features (random walk return probabilities, clustering coefficients) rather than in the number field redirects attention toward designing richer topological augmentations within the standard real-valued framework.

Future work could explore learned (rather than fixed) complex-valued initialization, complex-valued attention mechanisms, or non-harmonic initialization strategies that maintain phase diversity on regular graphs. However, our results suggest that the burden of proof for complex-valued approaches is high: any proposed method must demonstrate gains *beyond* what the same topological features provide in the real domain.

References

- Ralph Abboud, İsmail İlkan Ceylan, Martin Grohe, and Thomas Lukasiewicz. The surprising power of graph neural networks with random node initialization. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 2112–2118. ijcai.org, 2021. doi: 10.24963/IJCAI.2021/291.
- Linus Bao, Emily Jin, Michael M. Bronstein, İsmail İlkan Ceylan, and Matthias Lanzinger. Homomorphism counts as structural encodings for graph learning. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. URL <https://openreview.net/forum?id=qFw2RFJS5g>.
- Beatrice Bevilacqua, Fabrizio Frasca, Derek Lim, Balasubramaniam Srinivasan, Chen Cai, Gopinath Balamurugan, Michael M. Bronstein, and Haggai Maron. Equivariant subgraph aggregation networks. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL <https://openreview.net/forum?id=dFbKQaRk15w>.
- Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Liò, Guido F. Montúfar, and Michael M. Bronstein. Weisfeiler and lehman go cellular: CW networks. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 2625–2640, 2021.
- Giorgos Bouritsas, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M. Bronstein. Improving graph neural network expressivity via subgraph isomorphism counting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1):657–668, 2023. doi: 10.1109/TPAMI.2022.3154319.
- Jin-yi Cai, Martin Fürer, and Neil Immerman. An optimal lower bound on the number of variables for graph identification. *Comb.*, 12(4):389–410, 1992. doi: 10.1007/BF01305232.
- Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *J. Mach. Learn. Res.*, 24:43:1–43:48, 2023.
- Moshe Eliasof, Fabrizio Frasca, Beatrice Bevilacqua, Eran Treister, Gal Chechik, and Haggai Maron. Graph positional encoding via random feature propagation. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 9202–9223. PMLR, 2023.
- Fabrizio Frasca, Beatrice Bevilacqua, Michael M. Bronstein, and Haggai Maron. Understanding and extending subgraph gnns by rethinking their symmetries. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.
- Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 2017. URL <http://proceedings.mlr.press/v70/gilmer17a.html>.
- William L. Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural*

Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, pages 1024–1034, 2017.

Zhongyu Huang, Yingheng Wang, Chaozhuo Li, and Huiguang He. Going deeper into permutation-sensitive graph neural networks. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 9377–9409. PMLR, 2022.

Nicolas Keriven and Gabriel Peyré. Universal invariant and equivariant graph neural networks. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 7090–7099, 2019.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL <https://openreview.net/forum?id=SJU4ayYg1>.

Longlong Li, Cunquan Qu, and Guanghui Wang. Beyond node attention: Multi-scale harmonic encoding for feature-wise graph message passing. *CoRR*, abs/2505.15015, 2025. doi: 10.48550/ARXIV.2505.15015.

Derek Lim, Joshua David Robinson, Lingxiao Zhao, Tess E. Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL <https://openreview.net/forum?id=Q-UHqMorzil>.

Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL <https://openreview.net/forum?id=Syx72jC9tm>.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 4602–4609. AAAI Press, 2019. doi: 10.1609/AAAI.V33I01.33014602.

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

Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak A. Rao, and Bruno Ribeiro. Relational pooling for graph representations. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 4663–4673. PMLR, 2019. URL <http://proceedings.mlr.press/v97/murphy19a.html>.

Chendi Qian, Andrei Manolache, Christopher Morris, and Mathias Niepert. Probabilistic graph rewiring via virtual nodes. In *Advances in Neural Information Processing Systems 38: Annual*

Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024, 2024.

Ryoma Sato, Makoto Yamada, and Hisashi Kashima. Random features strengthen graph neural networks. In *Proceedings of the 2021 SIAM International Conference on Data Mining, SDM 2021, Virtual Event, April 29 - May 1, 2021*, pages 333–341. SIAM, 2021. doi: 10.1137/1.9781611976700.38.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL <https://openreview.net/forum?id=rJXMpikCZ>.

Yanbo Wang and Muhan Zhang. An empirical study of realized GNN expressiveness. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pages 52134–52155. PMLR, 2024.

Zehong Wang, Zheyuan Zhang, Tianyi Ma, Nitesh V. Chawla, Chuxu Zhang, and Yanfang Ye. Beyond message passing: Neural graph pattern machine. In *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, volume 267 of *Proceedings of Machine Learning Research*. PMLR, 2025.

Boris Weisfeiler and Andrei Leman. The reduction of a graph to canonical form and the algebra which appears therein. *NTI, Series 2*, 9:12–16, 1968.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL <https://openreview.net/forum?id=ryGs6ia5Km>.

Bohang Zhang, Lingxiao Zhao, and Haggai Maron. On the expressive power of spectral invariant graph neural networks. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pages 60496–60526. PMLR, 2024.

Xitong Zhang, Yixuan He, Nathan Brugnone, Michael Perlmutter, and Matthew J. Hirn. Magnet: A neural network for directed graphs. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 27003–27015, 2021.