Sparse Global Mixing via Compact-Support Graph Kernels

Himanshu Singh [1]

[1]*Computing and Artificial Intelligence Division (CAI-3), Los Alamos National Laboratory, Los Alamos, NM 87545, USA*. Correspondence to: Himanshu Singh singh_h@lanl.gov.

1. Introduction

Global pairwise mixing layers provide strong expressive power for graph representation learning, but incur quadratic interaction complexity and dense parameterization. This quadratic scaling limits interpret-ability and computational efficiency, particularly in global mixing architectures inspired by attention mechanisms.

We introduce a family of structured sparse global mixing operators based on compact-support graph kernels that explicitly restrict interactions to r-hop neighborhoods. Our proposed model, WendlandPairNet, leverages smooth compact-support kernels derived from shortest-path distances to induce geometrically grounded sparsity while preserving global aggregation capacity. WendlandPairNet derives its structured sparsity mask from compact-support Wendland kernels, whose finite support naturally induces sparse interaction matrices by restricting mixing to bounded-radius neighborhoods [1].

We evaluate our approach on standard graph classification benchmarks, including PROTEINS from the TU dataset collection [2], originally introduced in [3]. We compare:

1. DensePairNet — fully connected global mixing (conceptually related to global attention mechanisms [4, 5, 6]).
2. TopKPairNet — learned hard sparsification inspired by sparse attention approaches [7, 8].
3. **WendlandPairNet** — structured compact support sparsity.

Our experiments show that compact-support mixing reduces the effective interaction density to as low as $\sim 9\%$ of dense pairwise interactions (for $r = 3$) while maintaining competitive predictive accuracy. Unlike hard sparsification, which introduces instability and selection overhead, the compact-support design yields interpretable interaction patterns and a smooth trade-off between locality and expressivity.

Through controlled multi-seed and data-scarcity evaluations, we analyze the computational–statistical trade-offs induced by structured sparsity. Although dense mixing remains the strongest performer on PROTEINS, structured compact support substantially reduces effective interaction complexity with only modest degradation in predictive performance.

These findings suggest that compact-support operators provide a principled pathway toward scalable sparse global mixing mechanisms that balance interaction sparsity with expressive power, offering a geometrically grounded alternative to dense attention-style architectures.

2. Methods

Formally, we consider global mixing operators of the form given as follows:

$$h_i' = \sum_j \alpha_{i,j} W h_j,$$

where $\alpha_{i,j}$ is the *interaction mask*; W is the learnable linear transformation matrix applied to the node features before aggregation. Furthermore,

1. DensePairNet: $\alpha_{i,j} = 1$,
2. TopKPairNet: a hard sparse selection is applied,
3. WendlandPairNet: $\alpha_{i,j} = K\left(r^{-1} \cdot d(i,j)\right)$ through a compact-support Wendland kernel $K(\cdot)$.

3. Results

Table 1: Test Accuracy and Effective Interaction Density (10 Seeds) on PROTEINS.

Model	Interaction Density	Avg Interactions/Nodes	Test Acc.	Forward ms/batch
Dense	1.00	106.3	76.5 ± 0.0358	143.6
TopK	0.28	30	74 ± 0.0361	3613.5
Wend.	0.09	9.7	73.5 ± 0.0631	151.5

Figure 1 illustrates the accuracy–radius trade-off, demonstrating that increasing neighborhood radius improves approximation to dense global mixing.

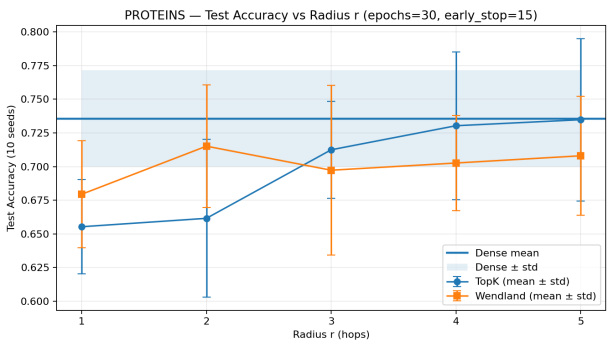

Fig. 1: Experimental verification of r-Sweep Tradeoff

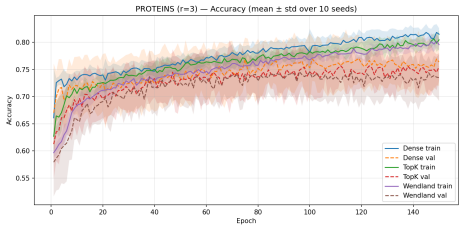

Fig. 3: Accuracy curves on PROTEINS (mean ± std over 10 seeds). WendlandPairNet achieves comparable generalization to DensePairNet while enforcing compact-support interaction sparsity, demonstrating effective structural regularization with significantly reduced interaction density.

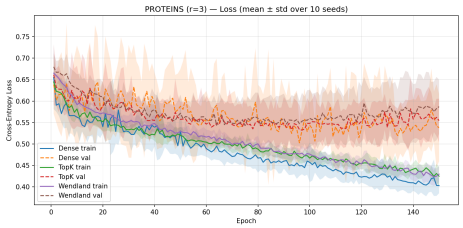

Fig. 4: Loss curves on PROTEINS (mean ± std over 10 seeds). WendlandPairNet achieves stable convergence and improved generalization by enforcing compact-support interaction, reducing overfitting compared to fully dense global mixing.

Notably, WendlandPairNet achieves nearly an order-of-magnitude reduction in interaction density relative to dense mixing while remaining within a small performance gap.

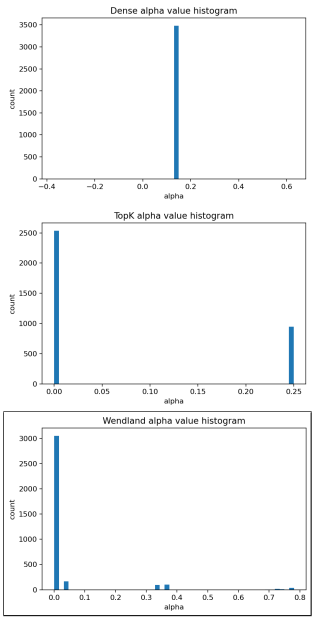

Fig. 2: Accuracy vs Interaction Density

4. Conclusion

We presented a structured sparse global mixing framework for graph neural networks based on compact-support graph kernels. By restricting pairwise interactions to geometrically defined r-hop neighborhoods, WendlandPairNet induces substantial structural sparsity while preserving competitive predictive performance relative to fully dense global mixing.

Our empirical analysis demonstrates a clear trade-off between interaction density and expressivity: compact-support operators reduce effective pairwise interactions by nearly an order of magnitude while maintaining strong accuracy on benchmark graph classification tasks. Unlike learned hard sparsification, the proposed structured approach yields stable training dynamics and interpretable interaction patterns.

These results highlight compact-support sparsity as a principled architectural alternative to dense attention-style mixing, offering a scalable pathway for global graph modeling that balances computational efficiency with expressive power. Future work will focus on sparse pair enumeration for realized runtime gains and on formalizing the capacity–locality trade-offs induced by compact support.

References

[1] Holger Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.

[2] 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.

[3] Karsten M Borgwardt, Cheng Soon Ong, Stefan Schönauer, SVN Vishwanathan, Alex J Smola, and Hans-Peter Kriegel. Protein function prediction via graph kernels. *Bioinformatics*, 21(suppl_1):i47–i56, 2005.

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

[5] Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. In *AAAI*, 2021.

[6] Chengxuan Ying et al. Do Transformers Really Perform Badly for Graph Representation? In *NeurIPS*, 2021.

[7] Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. Generating Long Sequences with Sparse Transformers. *arXiv preprint arXiv:1904.10509*, 2019.

[8] André F. T. Martins and Ramón Astudillo. From Softmax to Sparsemax. In *ICML*, 2016.