Multiresolution Graph Transformers and Wavelet Positional Encoding for Learning Long-Range and Hierarchical Structures

Nhat Khang Ngo * [1], Truong Son Hy * † [2] and Risi Kondor [3]

FPT Software AI Center [1], University of California San Diego [2], University of Chicago [3]
Co-first author *, Correspondent to tshy@ucsd.edu †

Summary

- We propose **Multiresolution Graph Transformers** (MGT), the first graph transformer architecture that can learn to represent large molecules at multiple scales, by utilizing a learning-to-cluster algorithm.
- We introduce **Wavelet Positional Encoding** (WavePE), a new positional encoding method that guarantees localization in spectral and spatial domains.
- We achieve competitive results on three macromolecule datasets consisting of polymers, peptides and protein-ligand complexes, and one drug-like molecule dataset. Significantly, our model outperforms other state-of-the-art methods and achieves chemical accuracy in estimating molecular properties (e.g., GAP, HOMO, and LUMO) calculated by **Density Functional Theory** (DFT) in the polymers dataset.
- Our PyTorch implementation is publicly available at:

`https://github.com/HySonLab/Multires-Graph-Transformer`

Motivation

- Macromolecules are **long-range** and **hierarchical** structures as they consist of many substructures.
- Substructures such as repeating units and functional groups are intrinsic parts of macromolecules; they present unique chemical reactions regardless of other compositions in the same molecules [4].

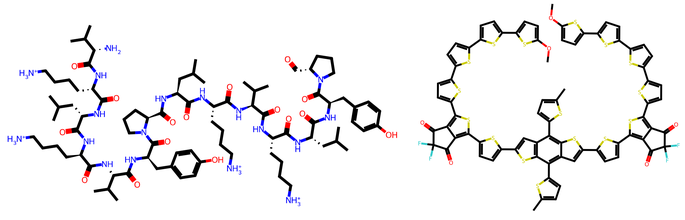


Figure 1: Examples of two macromolecules. a) An example of a peptide that consists of many functional groups. b) An example of a polymer that consists of many repeating units

Spectral Graph Wavelets

Employing graph wavelet signals allows capturing the relative positions of the nodes on graphs at multiple scales.

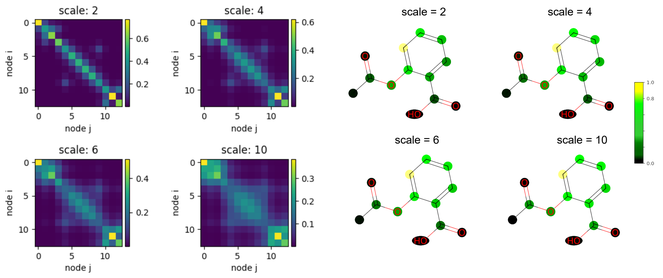
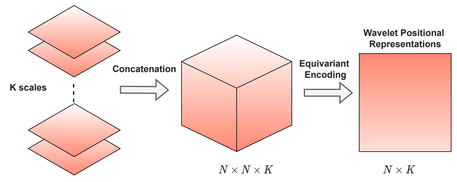


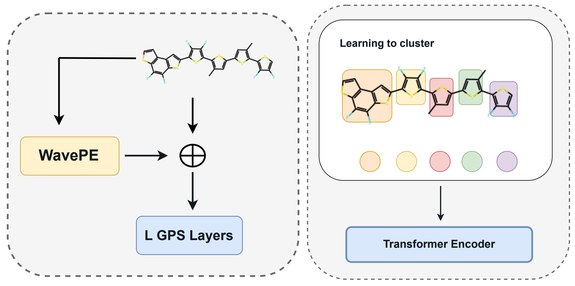
Figure 2: Visualization of some of the wavelets with scaling parameters on the Aspirin C9H8O4 molecular graph with 13 nodes (i.e. heavy atoms). The center node is colored yellow. The colors varying from bright to dark illustrate the diffusion rate from the center node to the others, i.e. nodes that are closer to the center node have brighter colors. Low-scale wavelets are highly localized, whereas the high-scale wavelets can spread out more nodes on the molecular graph

Wavelet Equivariant Encoding



- In general, we stack k wavelet matrices to obtain a $n \times n \times k$ wavelet tensors, where n is the graph's size.
- We employ high-order message passing networks [3, 5] to contract the tensor into a matrix of positional features with a size of $n \times k$.

Multiresolution Graph Transformer



MGT involves three primary components:

- We utilize GPS [6] a powerful graph transformer network to learn the **local** and **global** interactions of the atoms.
- A **learning-to-cluster** network learns to cluster the atom nodes into their meaningful substructures based on differentiable graph pooling [2, 9].
- Finally, a conventional transformer-like architecture is used to compute the representations of learnable substructures.

Peptides property prediction

Model	Params	Peptides-struct	Peptides-func
		MAE ↓	AP ↑
GCN	508k	0.3496 ± 0.0013	0.5930 ± 0.0023
GINE	476k	0.3547 ± 0.0045	0.5498 ± 0.0079
GatedGCN	509k	0.3420 ± 0.0013	0.5864 ± 0.0077
GatedGCN + RWPE	506k	0.3357 ± 0.0006	0.6069 ± 0.0035
Transformer + LapPE	488k	0.2529 ± 0.0016	0.6326 ± 0.0126
GPS	—	0.2500 ± 0.0005	0.6535 ± 0.0041
SAN + LapPE	493k	0.2683 ± 0.0043	0.6384 ± 0.0121
SAN + RWPE	500k	0.2545 ± 0.0012	0.6562 ± 0.0075
MGT + LapPE (ours)	499k	0.2488 ± 0.0014	0.6728 ± 0.0152
MGT + RWPE (ours)	499k	0.2496 ± 0.0009	0.6709 ± 0.0083
MGT + WavePE (ours)	499k	**0.2453 ± 0.0025**	**0.6817 ± 0.0064**

Table 1: Results on peptides property prediction. The datasets are included in [1].

Polymer property prediction

Model	Params	Property		
		GAP	HOMO	LUMO
DFT error		1.2	2.0	2.6
Chemical accuracy		0.043	0.043	0.043
GCN	527k	0.1094 ± 0.0020	0.0648 ± 0.0005	0.0864 ± 0.0014
GCN + Virtual Node	557k	0.0589 ± 0.0004	0.0458 ± 0.0007	0.0482 ± 0.0010
GINE	527k	0.1018 ± 0.0026	0.0749 ± 0.0042	0.0764 ± 0.0028
GINE + Virtual Node	557k	0.0870 ± 0.0040	0.0565 ± 0.0050	0.0524 ± 0.0010
GPS	600k	0.0467 ± 0.0010	0.0322 ± 0.0020	0.0385 ± 0.0006
Transformer + LapPE	700k	0.2949 ± 0.0481	0.1200 ± 0.0206	0.1547 ± 0.0127
MGT + LapPE (ours)	499k	**0.0378 ± 0.0004**	**0.0270 ± 0.0010**	0.0300 ± 0.0006
MGT + RWPE (ours)	499k	0.0384 ± 0.0015	0.0274 ± 0.0005	**0.0290 ± 0.0007**
MGT + WavePE (ours)	499k	0.0387 ± 0.0011	0.0283 ± 0.0004	0.0290 ± 0.0010

Table 2: Experimental results on the polymer property prediction task [7]. All the methods are trained in four different random seeds and evaluated by MAE ↓. Our methods are able to attain better performance across three DFT properties of polymers while having less number of parameters. All the properties are measured in eV.

Protein-ligand binding affinity prediction

Proteins are large and complex macromolecules that comprise one or more long chains of amino acid residues. Understanding the multiscale structure of proteins is important in estimating their fitness and functionality. In this experiment, we show the effectiveness of our model in capturing the long-range and hierarchical structures of proteins, that are larger than the peptides.

Method	3D-CNN	GNN	ENN	GVP-GNN	MGT + WavePE
RMSE ↓	**1.416 ± 0.021**	1.570 ± 0.025	1.568 ± 0.012	1.594 ± 0.073	1.436 ± 0.066

Table 3: Experimental Results on protein-ligand binding affinity dataset of ATOM3D benchmark [8].

References

References

[1] Vijay Prakash Dwivedi et al. "Long Range Graph Benchmark". In: *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. 2022. URL: https://openreview.net/forum?id=in7XC5RcjEn.

[2] Truong Son Hy and Risi Kondor. "Multiresolution equivariant graph variational autoencoder". In: *Machine Learning: Science and Technology* 4.1 (Mar. 2023), p. 015031. DOI: 10.1088/2632-2153/acc0d8. URL: https://dx.doi.org/10.1088/2632-2153/acc0d8.

[3] Truong Son Hy et al. "Predicting molecular properties with covariant compositional networks". In: *The Journal of chemical physics* 148.24 (2018).

[4] March Jerry. *Advanced Organic Chemistry: reactions, mechanisms and structure*. 1992.

[5] Haggai Maron et al. "Invariant and Equivariant Graph Networks". In: *International Conference on Learning Representations*. 2019. URL: https://openreview.net/forum?id=Syx72jC9tm.

[6] Ladislav Rampášek et al. "Recipe for a General, Powerful, Scalable Graph Transformer". In: *arXiv:2205.12454* (2022).

[7] Peter C St. John et al. "Message-passing neural networks for high-throughput polymer screening". In: *The Journal of chemical physics* 150.23 (2019), p. 234111.

[8] Raphael John Lamarre Townshend et al. "ATOM3D: Tasks on Molecules in Three Dimensions". In: *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*. 2021. URL: https://openreview.net/forum?id=FkDZLpK1Ml2.

[9] Rex Ying et al. "Hierarchical Graph Representation Learning with Differentiable Pooling". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS'18. Montréal, Canada: Curran Associates Inc., 2018, pp. 4805–4815.