Topological Encoding in Transformer for Molecular Representation

Jiahao Lai

2023.12.14

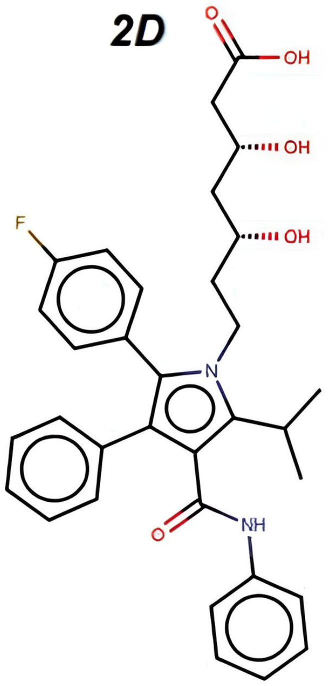
Background

Molecules embody multi-modal representations, encompassing 1-D SMILES sequences, 2-D graph representations, and 3-D conformations.

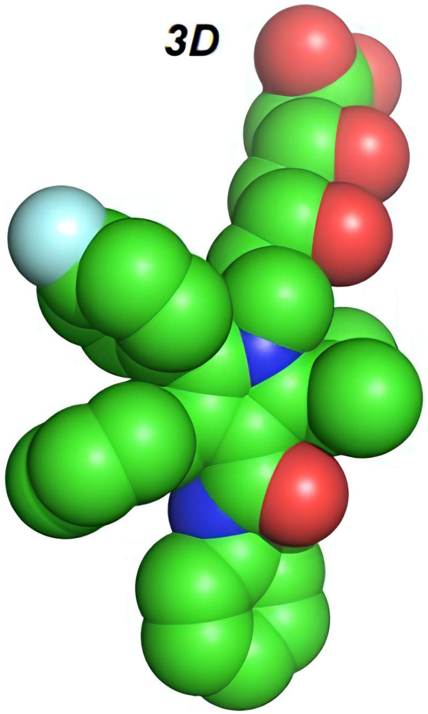
0D/1D

$C_{33}H_{35}FN_2O_5$

2D



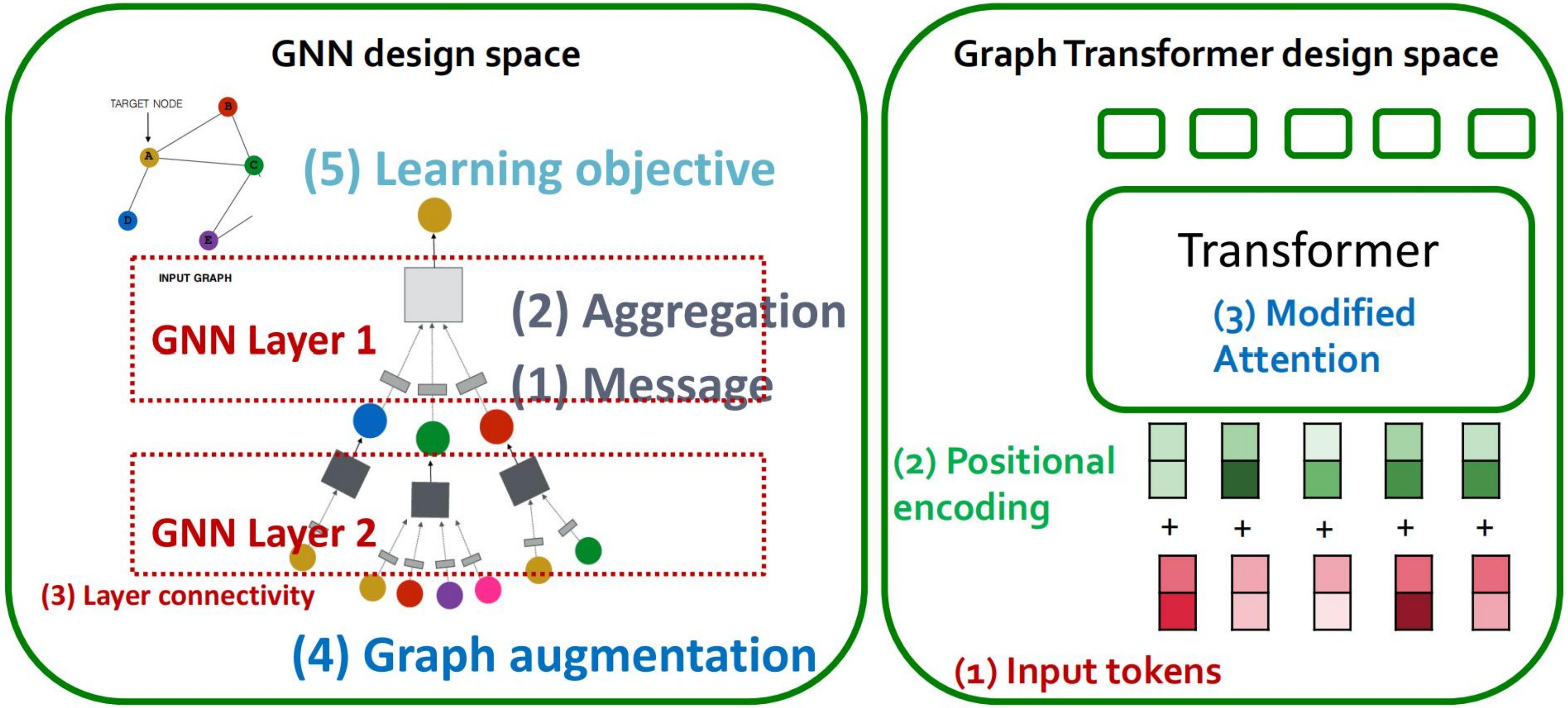
3D



https://drzinph.com/computing-molecular-descriptors-intro/?preview=true&_thumbnail_id=206

Background

Graph neural networks (GNNs) and Transformers stand as two primary architectural categories for molecular representations.



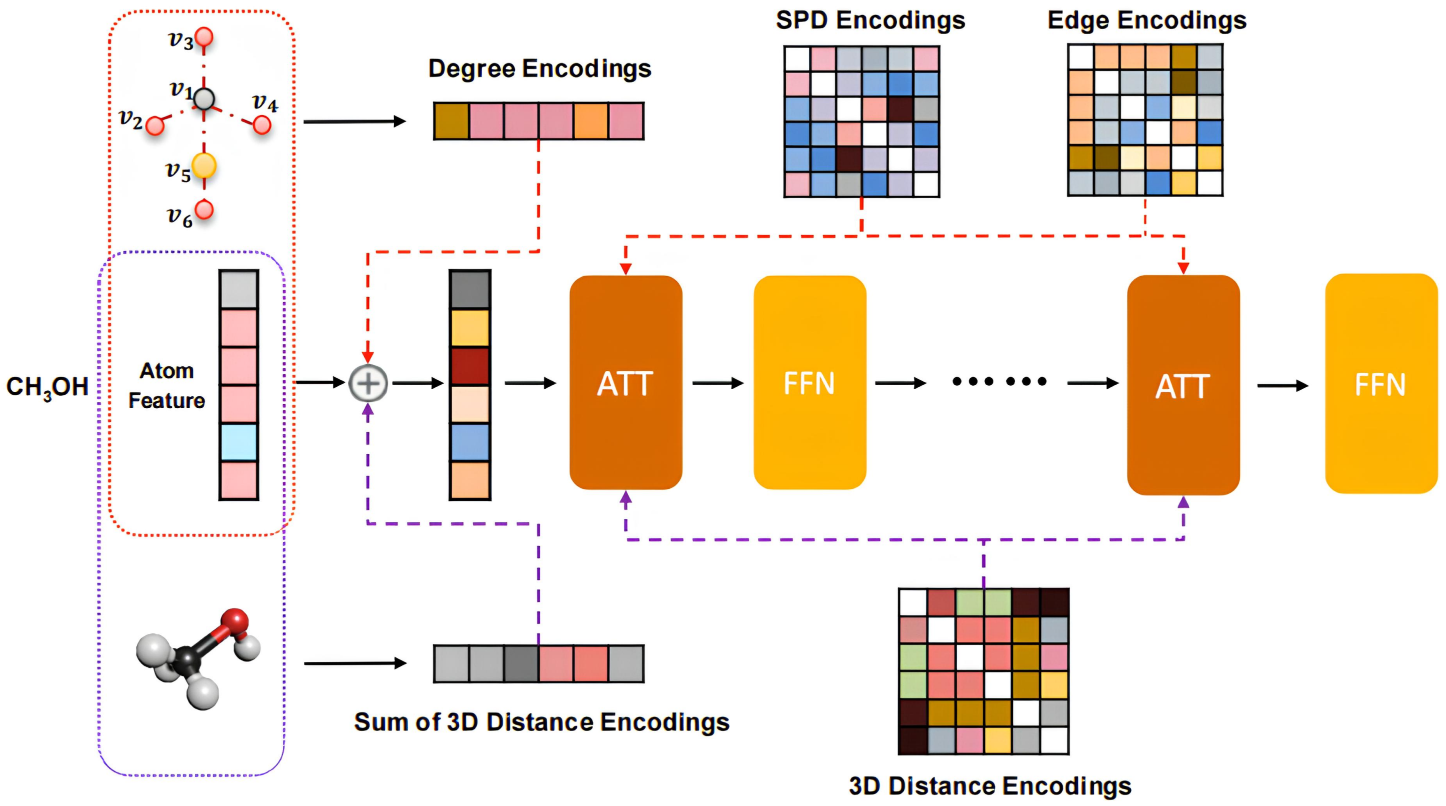
Jure Leskovec, S 2023, Module 1: Machine Learning with Graphs, CS224W, Lecture notes, Stanford University

Motivation and Research Question

- GNNs face difficulties in capturing long-range dependencies because graph convolutions are designed for local feature aggregation, and deep GNNs encounter challenges related to over-smoothing.
- Transformers are not suitable to model molecules since self-attention neglects inherent graph structure and global correlation disrupts locality between nodes.
- How can molecular multi-modal information be effectively integrated into Transformer models, particularly by leveraging their topological structures?

Related Work

Transformer-M(ICLR2023) encodes 2D and 3D structural information of molecules into Transformer module.

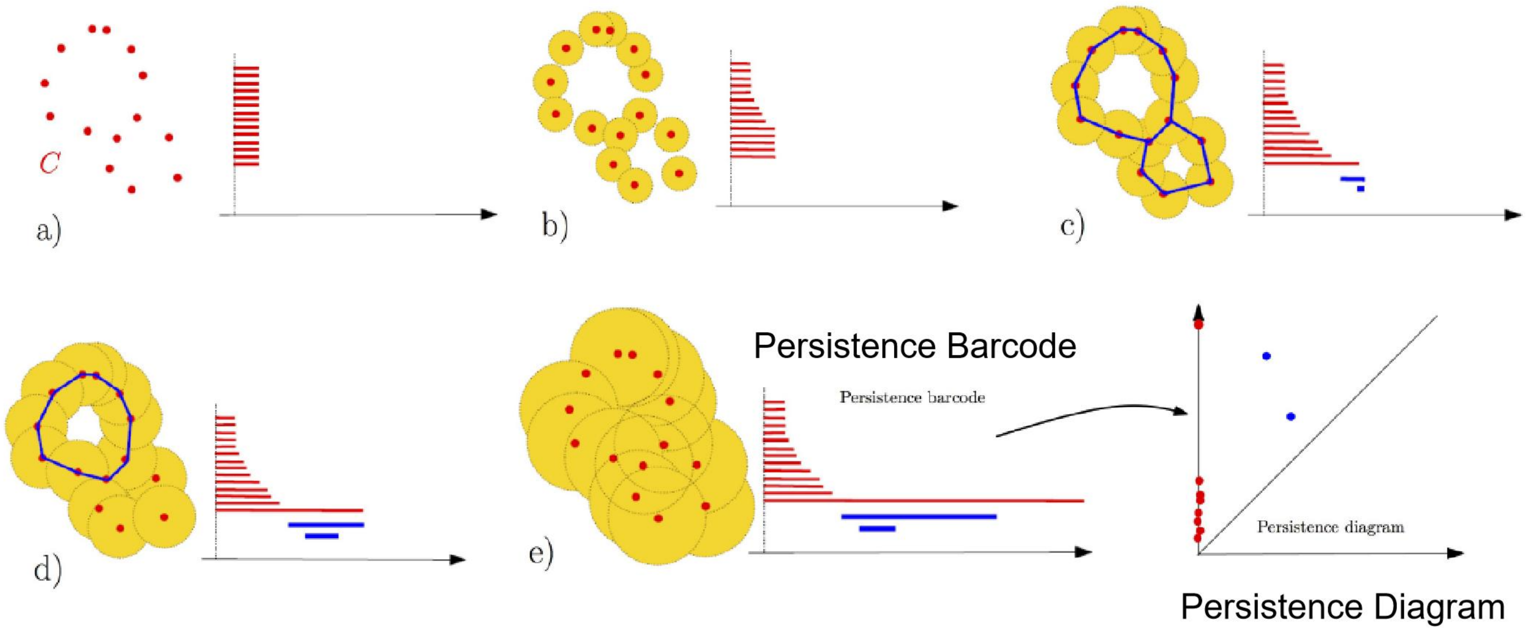


Luo, S., Chen, T., Xu, Y., Zheng, S., Liu, T. Y., Wang, L., & He, D. (2022). One transformer can understand both 2d & 3d molecular data. arXiv preprint arXiv:2210.01765.

Our Method

- Persistent Homology captures multi-scale topological features of molecular 3-D conformation.

- We encode Persistent Homology-based information into Transformer models.



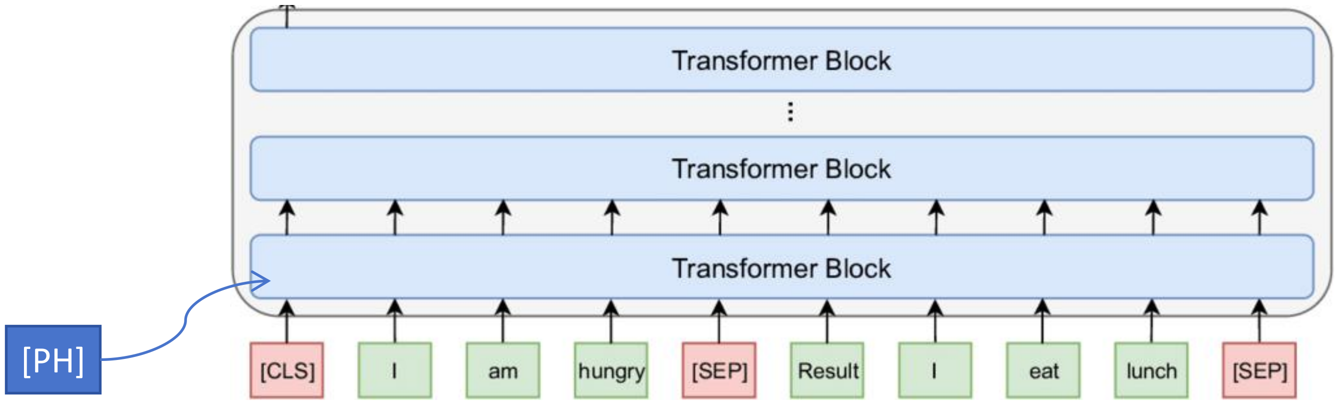
Chazal, F., & Michel, B. (2021). An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. Frontiers in Artificial Intelligence, 4, 667963.

Topological Global Encoding

- We transform Persistence Diagrams into vector γ using different vectorization methods.

- Inspired by BERT, we attach a special token [PH] with different vector value γ at the beginning of each sequence in multi-heads, to represent the graph-level feature.

- To distinguish the correlation between node attributes and [PH], we parametrize all key-value-query encodings for [PH] distinctly.



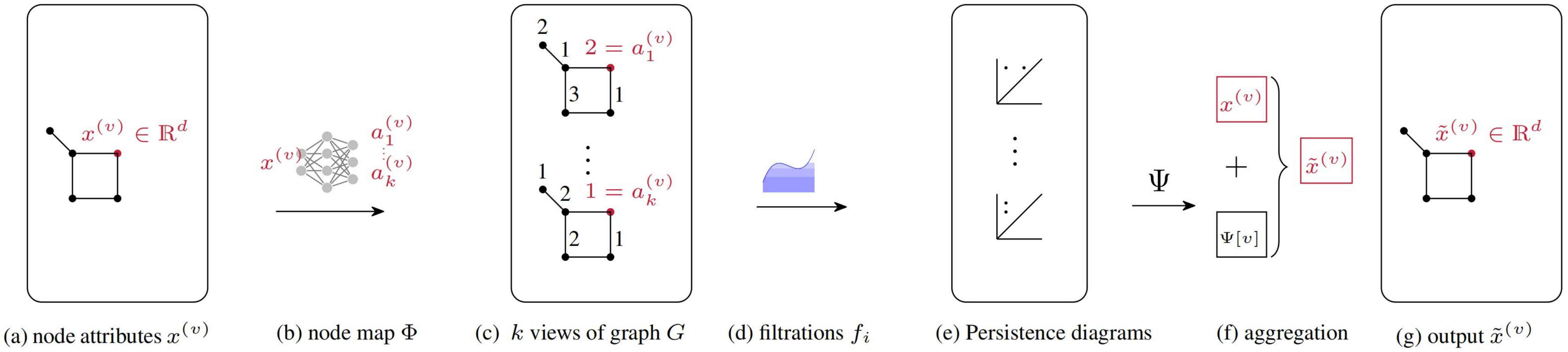
Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Topological Local Encoding

- Motivated by TOGL(ICLR2022), we embed Persistence Diagrams into a high-dimensional space that will be used to obtain the node encodings.

- Embedding transformations are local in that they apply to each single point without taking the other points into account.

$$\Psi^{(l)}: \left\{\mathcal{D}_1^{(l)}, \ldots, \mathcal{D}_k^{(l)}\right\} \to \mathbb{R}^{n' \times d} \qquad \tilde{x}^{(v)} = x^{(v)} + \Psi^{(0)}\left(\mathcal{D}_1^{(0)}, \ldots, \mathcal{D}_k^{(0)}\right)[v]$$



Horn, M., De Brouwer, E., Moor, M., Moreau, Y., Rieck, B., & Borgwardt, K. (2021). Topological graph neural networks. arXiv preprint arXiv:2102.07835.

Conclusion and Future Work

- We propose a scheme to encode both global and local topological information of molecules into Transformer models.

- Global topological encoding categorizes molecules through persistent homology vectorization features, while local topological encoding emphesizes critical nodes corresponding to each barcode attributes.

- Future work
  1. Coding and experiment enhencement
  2. Theoretical formulation and analysis
  3. Substructural topological encoding