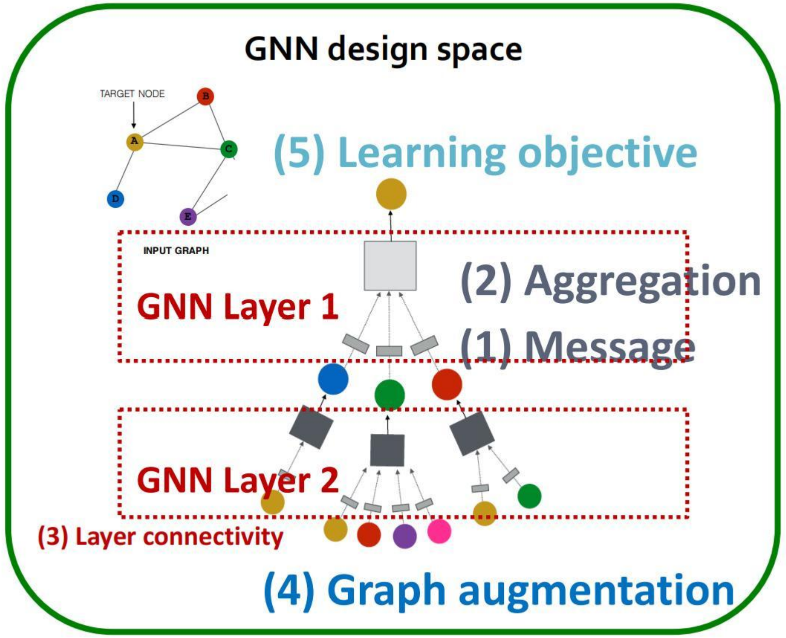
Message Passing Go Topological: High-Order Graph Transformers

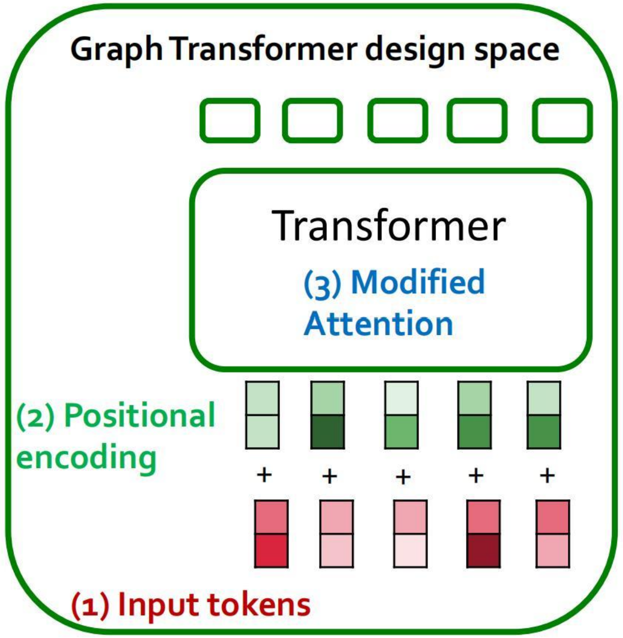
Jiahao Lai

10.23.2024

Graph Transformer Overview

- GNNs are biased towards encoding local structure and unable to capture global or long-range information (e.g. over-smoothing, over-squashing)
- Graph Transformers (GT) attend information over all nodes and hence are not limited to local aggregation.





$$\mathbf{h}_v^{(t+1)} := \sigma\left(\mathbf{h}_v^{(t)}\mathbf{W}_1^{(t)} + \sum_{w\in N(v)} \mathbf{h}_w^{(t)}\mathbf{W}_2^{(t)}\right),$$

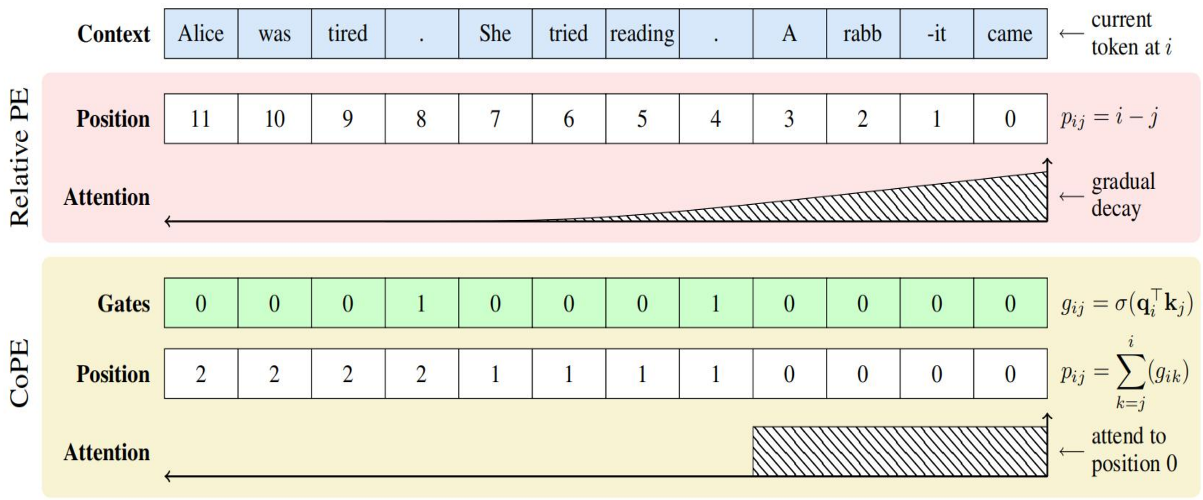
VS

$$\mathrm{Attn}(\mathbf{X}^{(t)}) := \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V},$$

Positional Encodings

- General GT architecture is less expressive in distinguishing non-isomorphic graphs than standard GNNs because attentions are permutation-invariant.
- Well-designed positional encodings (PE) are needed to capture graph structure.
- Analogous to language models: RoPE and contextual PE

$$\boldsymbol{\mathcal{R}}_{x,y} = \left(\begin{array}{cc|cc} \cos x\theta & -\sin x\theta & 0 & 0 \\ \sin x\theta & \cos x\theta & 0 & 0 \\ \hline 0 & 0 & \cos y\theta & -\sin y\theta \\ 0 & 0 & \sin y\theta & \cos y\theta \end{array}\right)$$



- Graph data itself doesn’t possess coordinate system.
- Positional encodings are supposed to align with contextual information on graphs.

Positional Encodings

Local PE (node, edge and subgraph):

- node degrees
- shortest-path distance of a node to a set of anchors
- eigenvectors of the adjacency or Laplacian
- random-walk kernels
- Ricci curvature
- shortest-path distances between two nodes

Global PE (entire graph):

- eigenvalues of the adjacency or Laplacian
- number of connected components
- diameter

Invariance (graph-level tasks) and equivarience (node-level tasks) for geometric features e.g. 3D coordinates of atoms, angles of bonds, or torsion angles of planes:

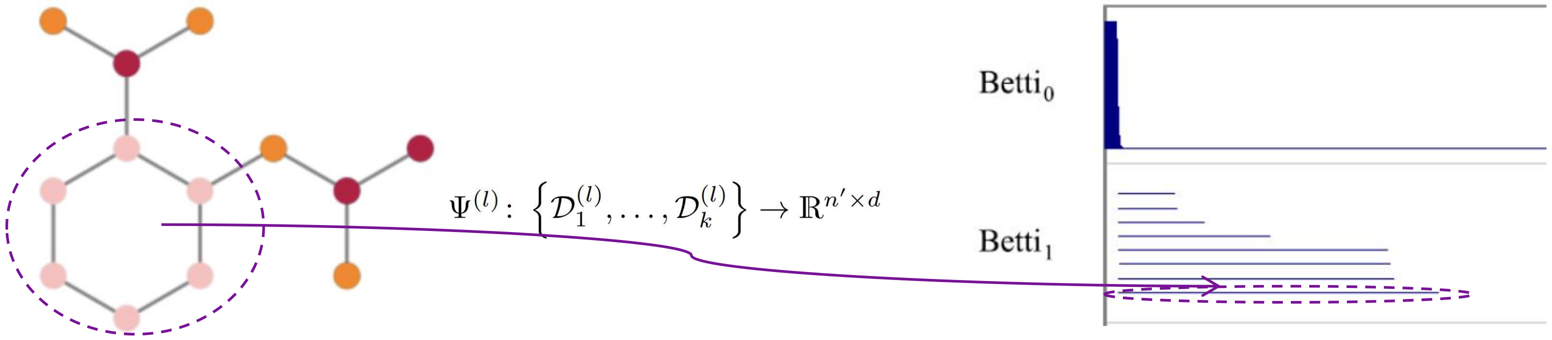
- Gaussian kernels to encode 3D distances between pairs.

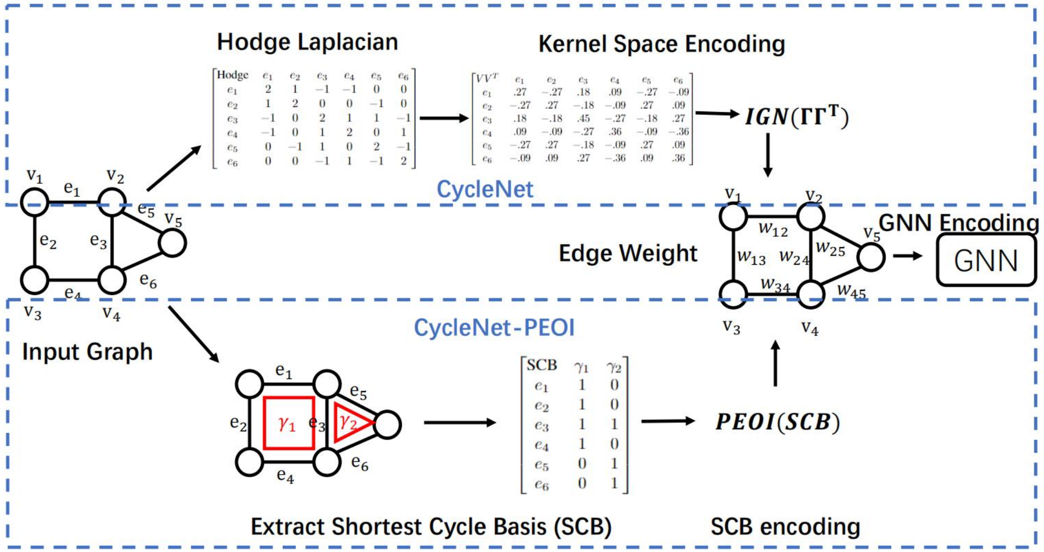
$$h_i^{(0)} = x_i + z^-_{\deg^-(v_i)} + z^+_{\deg^+(v_i)},$$

$$A_{ij} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b_{\phi(v_i,v_j)} + c_{ij},$$

Positional Encodings

- In particular, GNNs are limited in their capability of detecting graph structure such as triangles or cliques.



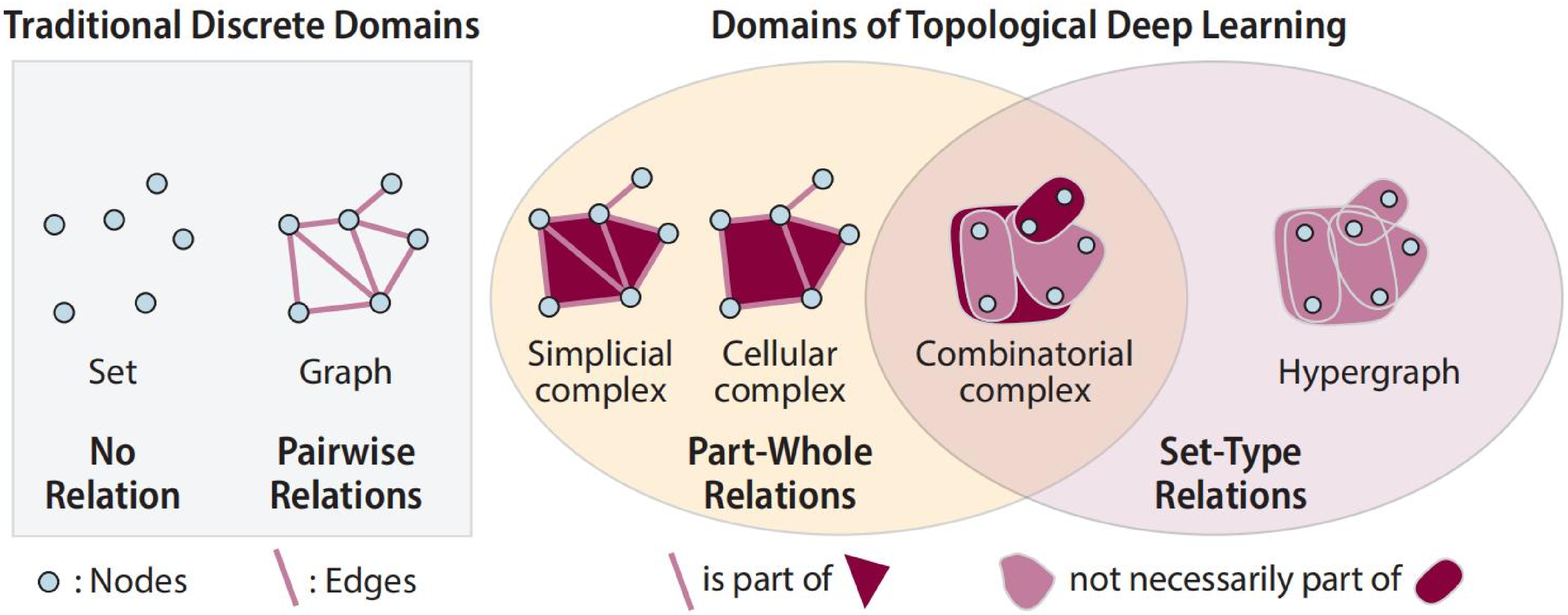


Message Passing

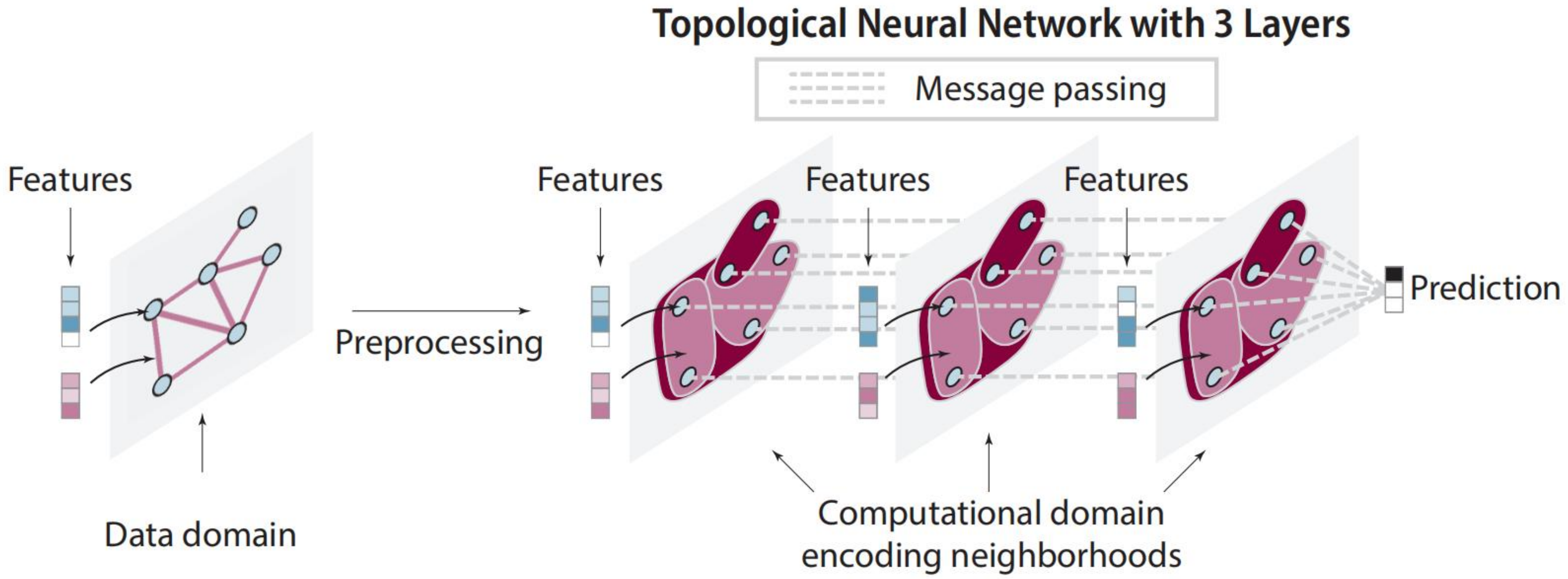
- Topological Deep Learning (TDL) defines a class of new domains with high-order relations and proposes a class of models that perform message passing on corresponding domains.

- While other architectures equivalent to higher-dimensional k-WL tests have been proposed, they suffer from high computational and memory complexity, and lack the key property of GNNs: Locality. TDL tackle this problem by considering local higher-order interactions.

Message Passing



Message Passing



Message Passing