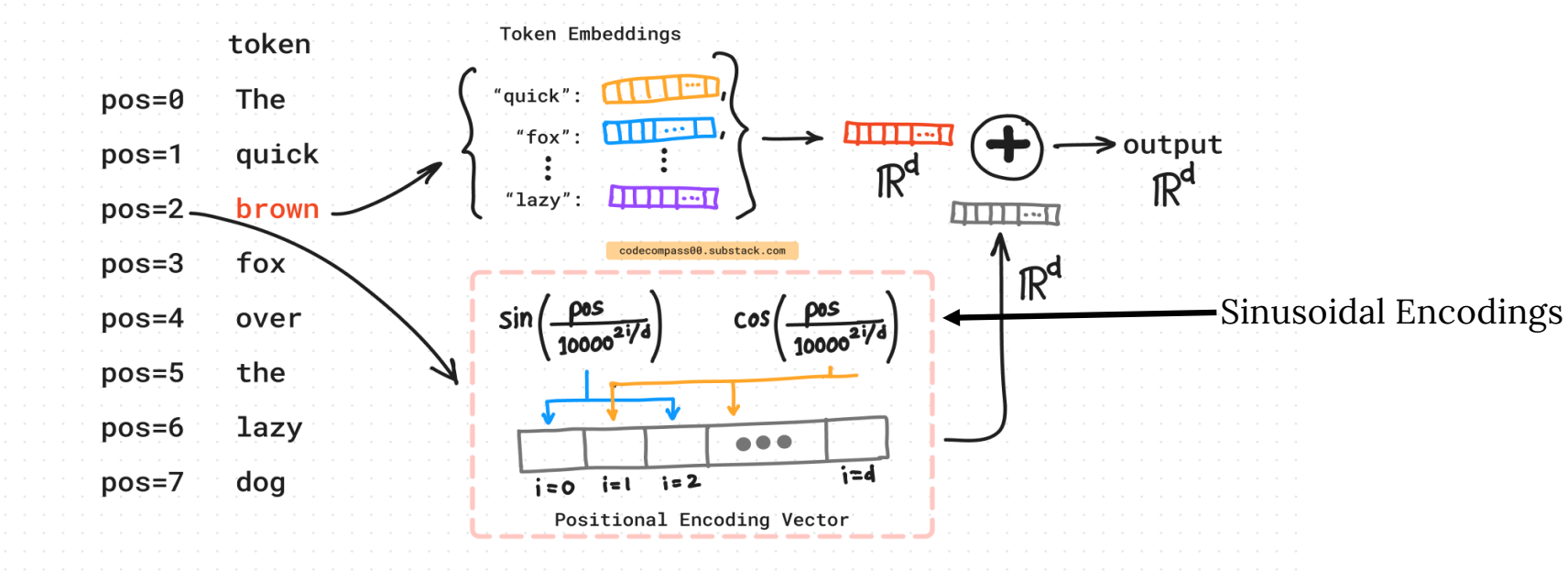
ETH zürich

Sign and Basis Invariant Networks for Spectral Graph Representation Learning

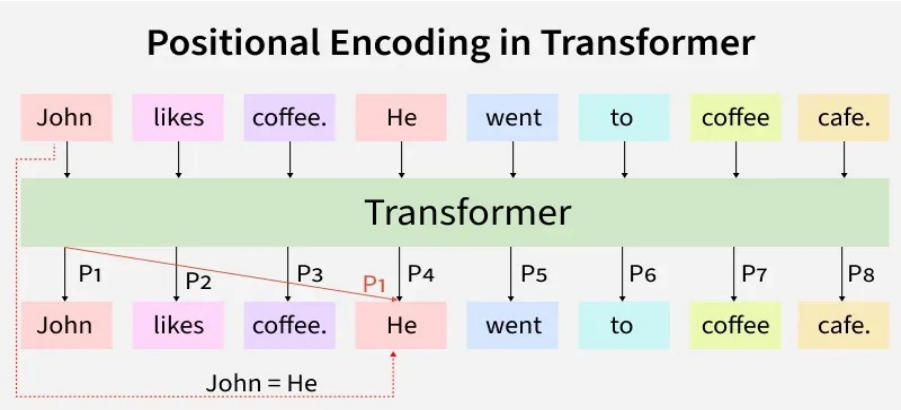
Prathamesh Tagore

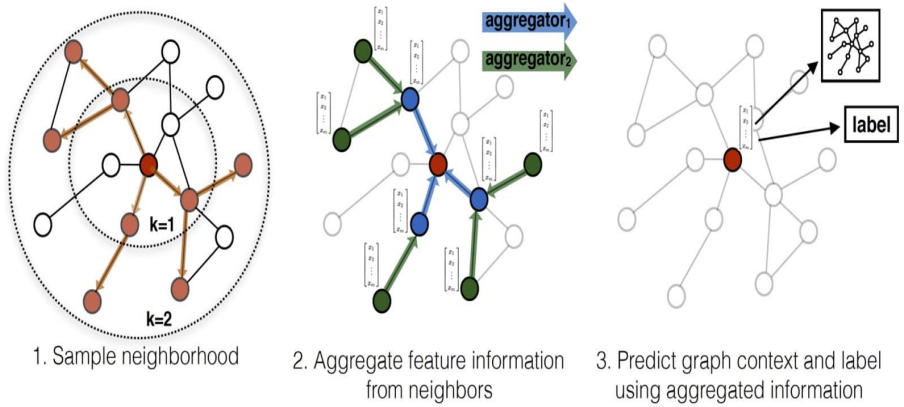
Positional Encodings (PE) – What are they?



- Tells you about the position of each piece of data in the larger structure
- Gives your model an idea about the overall topology

Positional Encodings (PE) – Why use them?





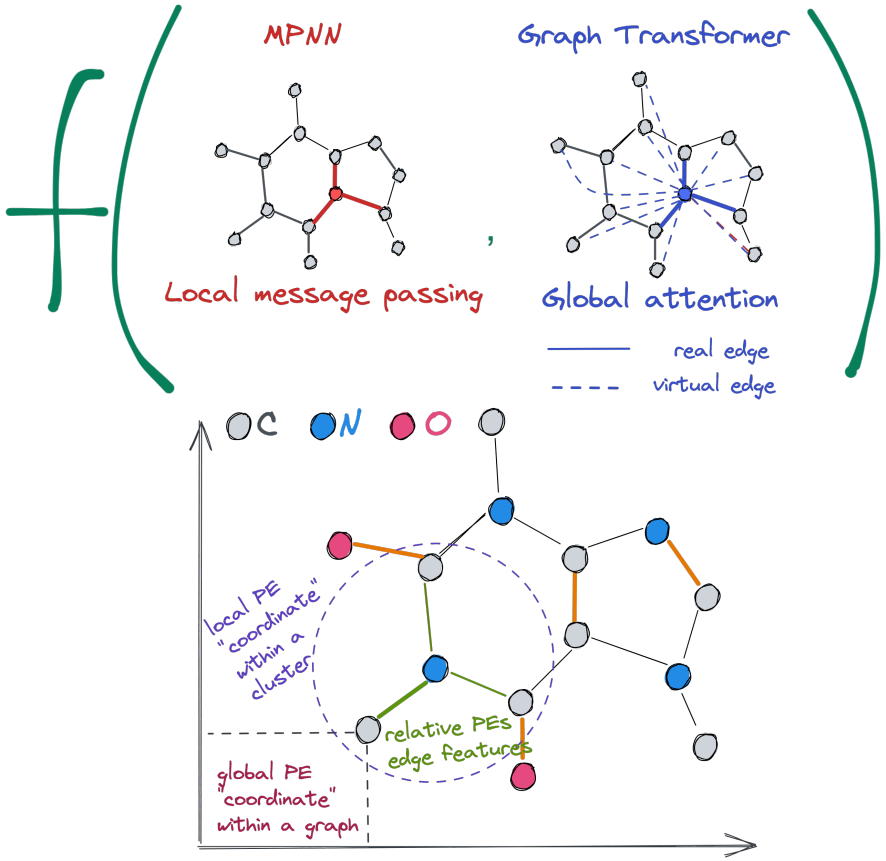
1. Sample neighborhood

2. Aggregate feature information from neighbors

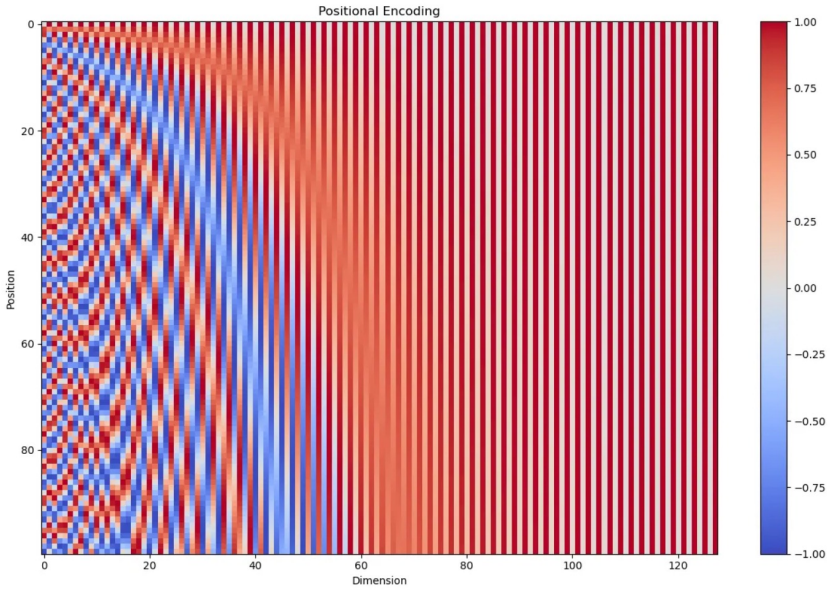
3. Predict graph context and label using aggregated information

- Required for establishing connections in data
- Helps break symmetry – improves capability to capture dependencies
- GNNs can have trouble differentiating non-isomorphic graphs – PEs help solve this

Positional Encodings (PE) – Why use them?



- Major applications include graph transformers and link prediction
- Also used for Graph Clustering and Partitioning

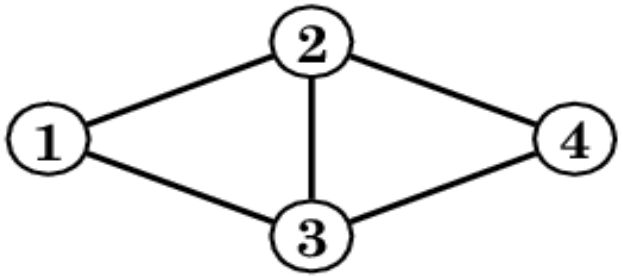


Graph Laplacian

$$L = D - A$$

1. D is the degree matrix which has the sum of all incoming edges for all graph nodes, A is the adjacency matrix
2. Helps in capturing distances, subgraph, local structures, etc.

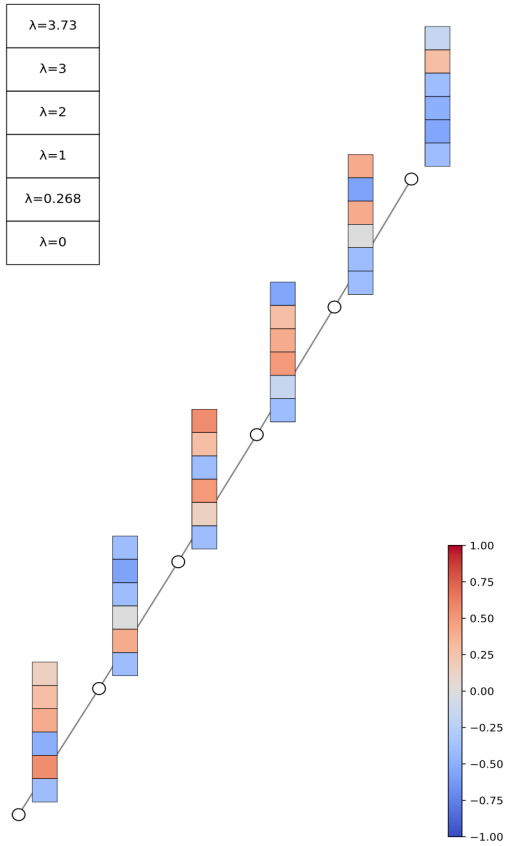
$$D = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$



$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

$$L = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}$$

Laplacian Positional Encodings (LapPEs)



- Compute k eigenvectors associated with the smallest non-zero eigenvalues
- For any node i, its LapPE is a vector containing the i-th component of these k eigenvectors
- Gives unique signatures to all nodes for capturing structural properties
- Low eigenvalues of Laplacian capture global clusters and graph skeleton
- High eigenvalues of Laplacian capture local details and high frequency noise

Laplacian Positional Encodings (LapPEs)

vec1	vec2	vec3	vec4
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Sorted per eigenvalues →

Eigenvectors Matrix

1
2
3

PE for node 1 (k = 3) – 1st element taken from first 3 eigenvectors

5
6
7

PE for node 2 (k = 3) – 2nd element taken from first 3 eigenvectors

Alternative Encodings: Heat kernel-based

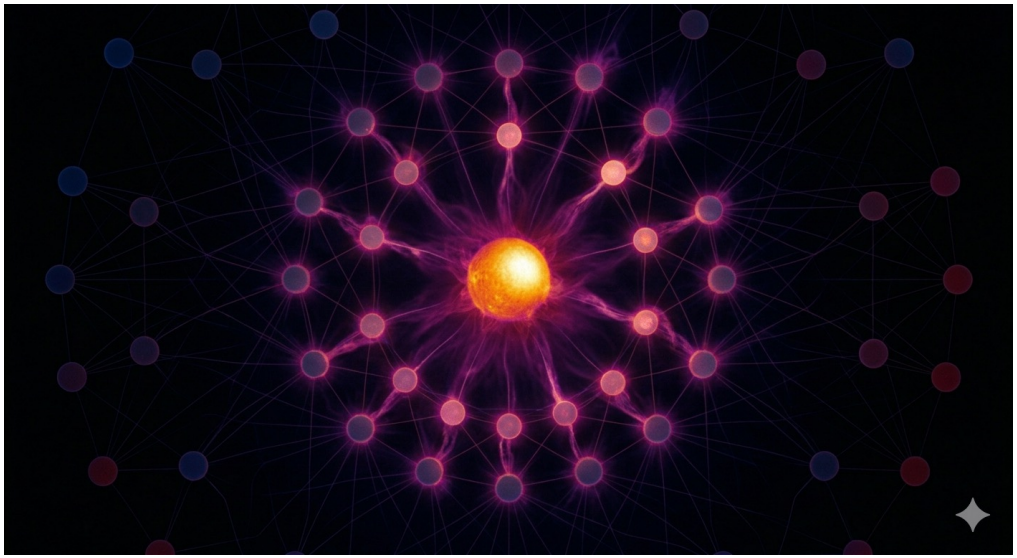


Image Credit: Gemini

- A graph node is the source of heat
- Heat dissipates from this node and reaches to other nodes
- Received heat for a node depends on its distance from the heat source node and time

Some alternatives for positional encodings

- Heat kernel-based PE
  - Compute the heat equation.
    $H_t = \exp(-tL)$ (L is laplacian, t is diffusion time – a hyperparameter)
  - $(H_t)_{ii} = \Sigma_{j=1}^{n} e^{-t\lambda_j}(i)^2$
  - Pick different set of times to create Heat Kernel PE

$$PE_{heat}(i) = [(H_{t1})_{ii}, (H_{t2})_{ii}, \ldots., (H_{tk})_{ii}]$$

- Random walk-based PE
  - Compute the random walk matrix.

$$P = D^{-1}A$$

  - Calculate landing probabilities. Get diagonal of the kth power of the transition matrix.

$$RW_i(k) = \left(P^k\right)_{ii}$$

  - Construct the encoding vector by taking a sequence of values (for different k) in step 2

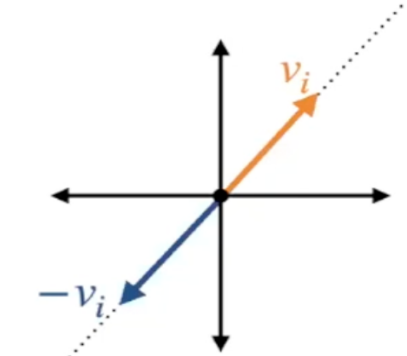
$$PE_{RW}(i) = [P_{ii}^1, P_{ii}^2, \ldots., P_{ii}^N]$$

- ....

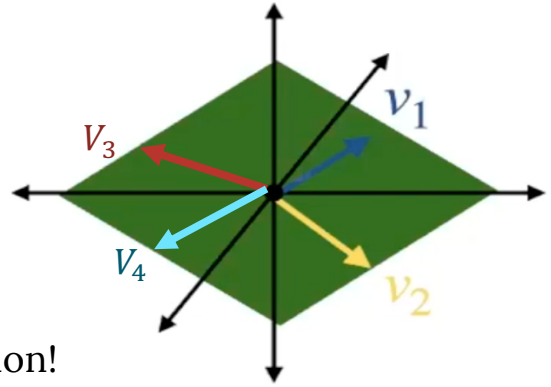
Eigenvector observations

- If v is an eigenvector, then so is −v. An eigen solver may return either.

$$Lv = \lambda v \implies L(-v) = \lambda(-v)$$



- Many bases describe the same d-dim eigenspace. Not generalized with existing work in this direction.



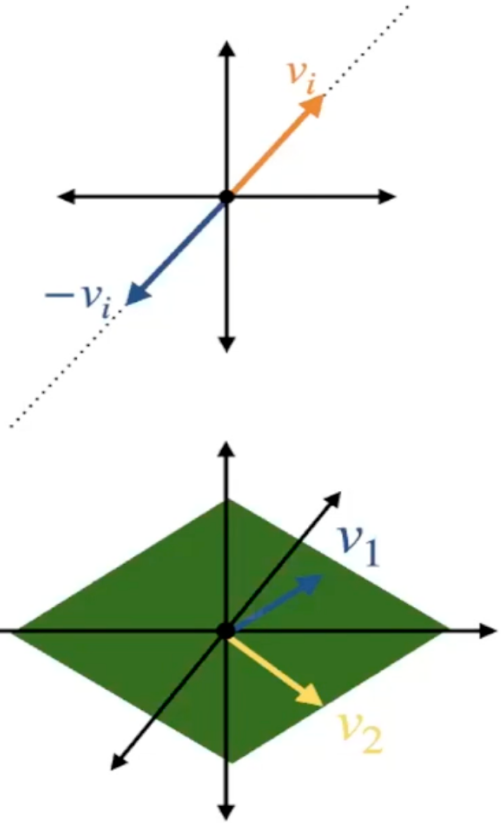
Considering for these observations is important for generalization!

V_1, V_2 *and* V_3, V_4 *are both basis vector pairs*

The Problem to Solve

- Sign Invariance: $f(v) = f(-v)$

- Basis Invariance: $f(V) = f(VQ)$ for all Q in orthogonal group O(d)

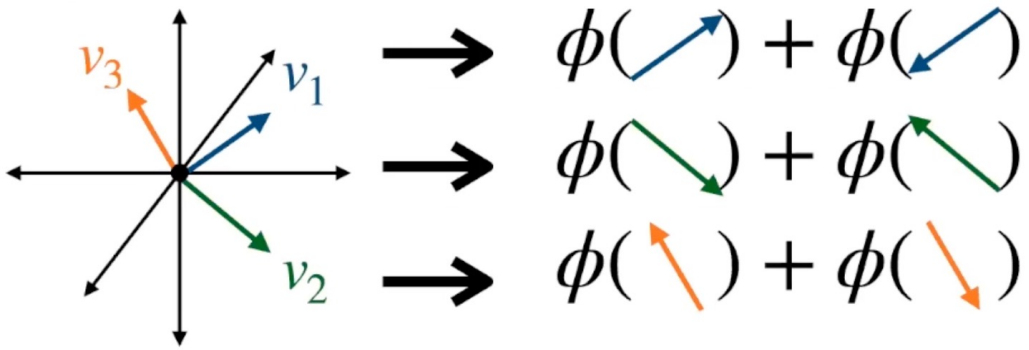


How to design Graph Neural Networks that can utilize Laplacian eigenvectors as positional encodings while remaining mathematically invariant to their inherent symmetries?

Sign Invariance in One Subspace

Characterization:

- f: $R^n \rightarrow R^n$ is continuous and sign invariant if and only if
  $f(v) = \emptyset(v) + \emptyset(-v)$ *for some continuous* ϕ



Universal Architecture:

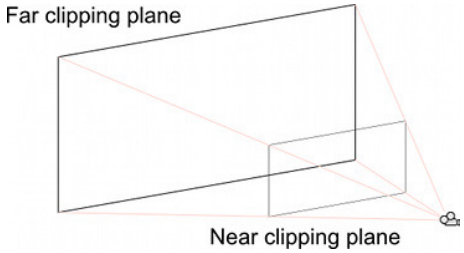
$f(v) = \emptyset(v) + \emptyset(-v)$

General f:

$\emptyset = MLP$

Projection Matrices and how are they affected by basis

- Projection matrix helps us realize how a region would project onto a different dimension
- It is independent of a specific basis



For the following, the columns of A and B are different sets of basis vectors for the same subspace

If columns of A, B are orthonormal (orthogonal to each other and length 1)

$$P_A = AA^T, P_B = BB^T$$
$$B = AQ$$
$$P_B = (AQ)(AQ)^T = A(QQ^T)A^T = P_A$$

For the general case,

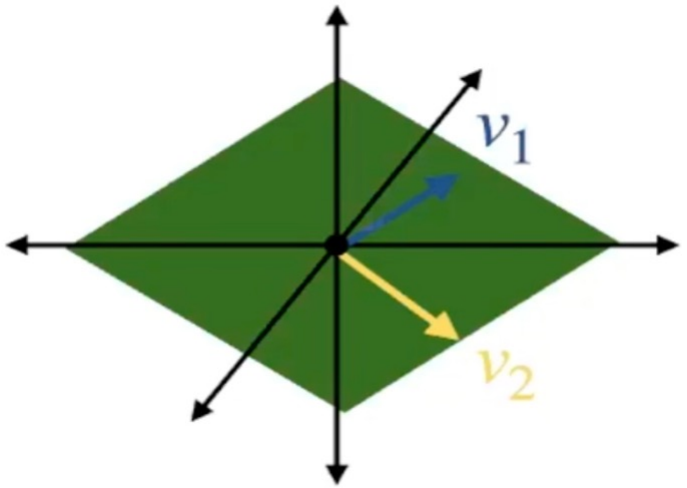
$$P_A = A(A^TA)^{-1}A^T, P_B = B(B^TB)^{-1}B^T$$
$$B = AQ$$
$$P_B = (AQ)\left((AQ)^T(AQ)\right)^{-1}(AQ)^T$$
$$P_B = AQ(Q^TA^TAQ)^{-1}Q^TA^T$$
$$P_B = AQ(Q^{-1}(A^TA)^{-1}(Q^T)^{-1})Q^TA^T$$
$$P_B = A(A^TA)^{-1}A^T = P_A$$

Basis Invariance in One Subspace

Characterization:

- f: $R^{nxd} \rightarrow R^n$ is continuous and $f(VQ) = f(V)$ for all orthogonal Q if and only if $f(V) = \emptyset(VV^T) for\ some\ continuous\ \emptyset$

$$(VQ)(VQ)^T = V(QQ^T)V^T = VV^T$$



Decomposing into Single Subspaces

Goal: $f(V_1 Q_1 \ \ldots., V_l Q_l) = f(V_1, \ldots., V_l), \quad (Q_1, \ldots., Q_l) \ \in O(d_1)\, x \ \ldots \ x\, O(d_l)$

Decomposition Theorem:

Under mild topological conditions, every continuous f invariant to $G = G_1\, x \ \ldots x\, G_l$ can be represented

$$f(x_1, \ldots., x_l) = \ \rho(\emptyset(x_1), \ldots., \emptyset(x_l)), \text{ and}$$

1. $\emptyset_i$ is G_i invariant
2. If $G_i = G_j, can\ take\ \emptyset_i = \ \emptyset_j$

- Takeaway: only need to do G_i invariance for $G_1\, x \ \ldots x\, G_l$ invariance!

SignNet: Sign Invariant Neural Net

Decomposition Theorem:

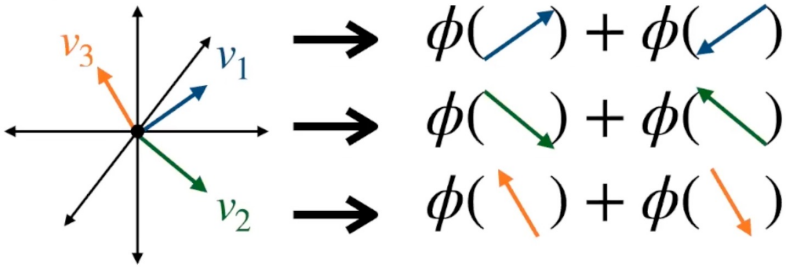
Under mild topological conditions, every continuous f invariant to $G = G_1 \, x \, \ldots x \, G_l$ can be represented

$$f(x_1, \ldots, x_l) = \rho(\emptyset(x_1), \ldots, \emptyset(x_l)), \text{ and}$$

1. $\emptyset_i$ is G_i invariant
2. If $G_i = G_j, can\ take\ \emptyset_i = \emptyset_j$

SignNet: $f(v_1, \ldots, v_l) = \rho(\emptyset(v_1) + \emptyset(-v_1), \ldots, \emptyset(v_l) + \emptyset(-v_l))$

- $\rho, \emptyset$: MLP, DeepSets, Transformers, or GNN



BasisNet: Basis Invariant Neural Net

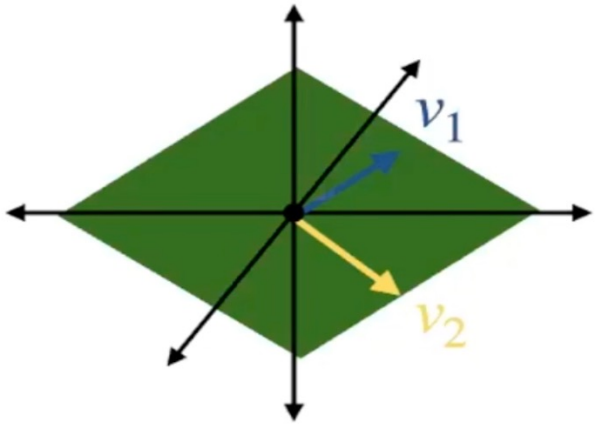
Decomposition Theorem:

Under mild topological conditions, every continuous f invariant to $G = G_1 \, x \, \ldots x \, G_l$ can be represented

$$f(x_1, \ldots, x_l) = \rho(\emptyset(x_1), \ldots, \emptyset(x_l)), \text{ and}$$

1. $\emptyset_i$ is G_i invariant
2. If $G_i = G_j, can\ take\ \emptyset_i = \emptyset_j$

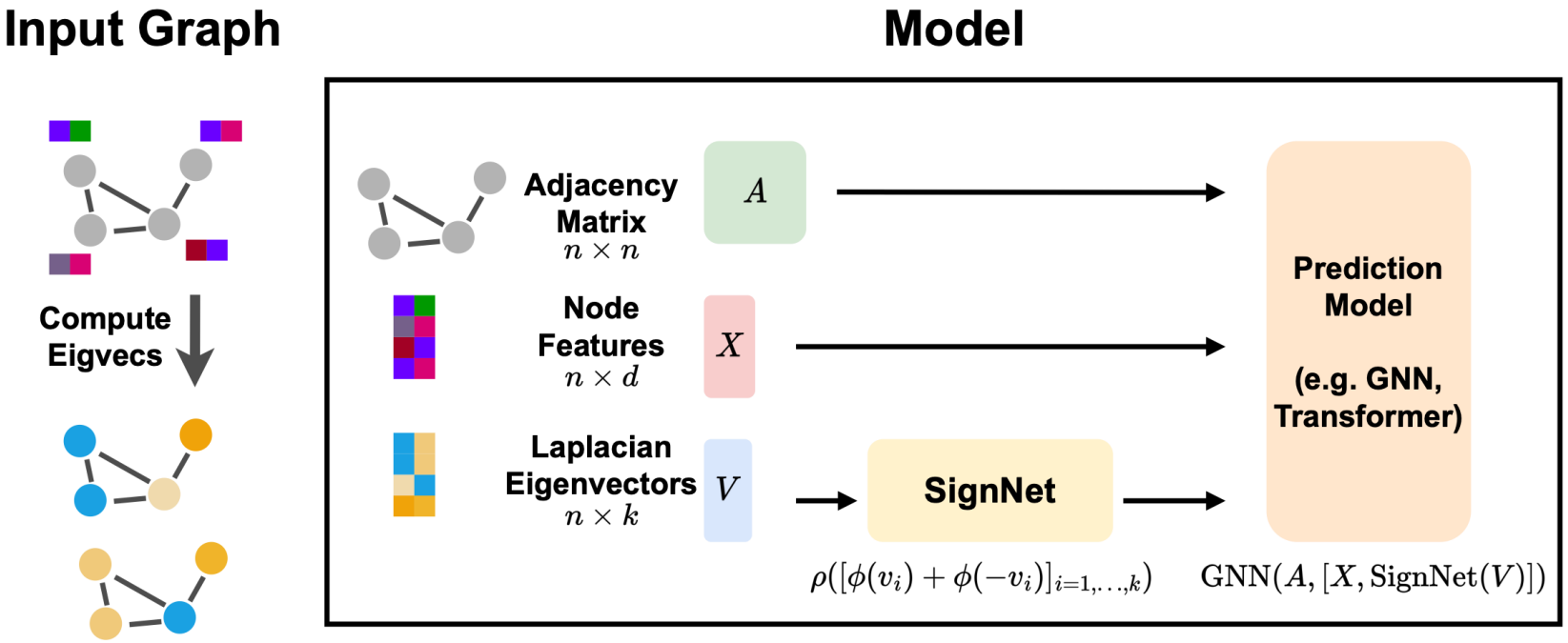
BasisNet: $f(V_1, \ldots, V_l) = \rho([\emptyset_{d_i}(V_i V_i^T)]_{i=1,\ldots,l})$



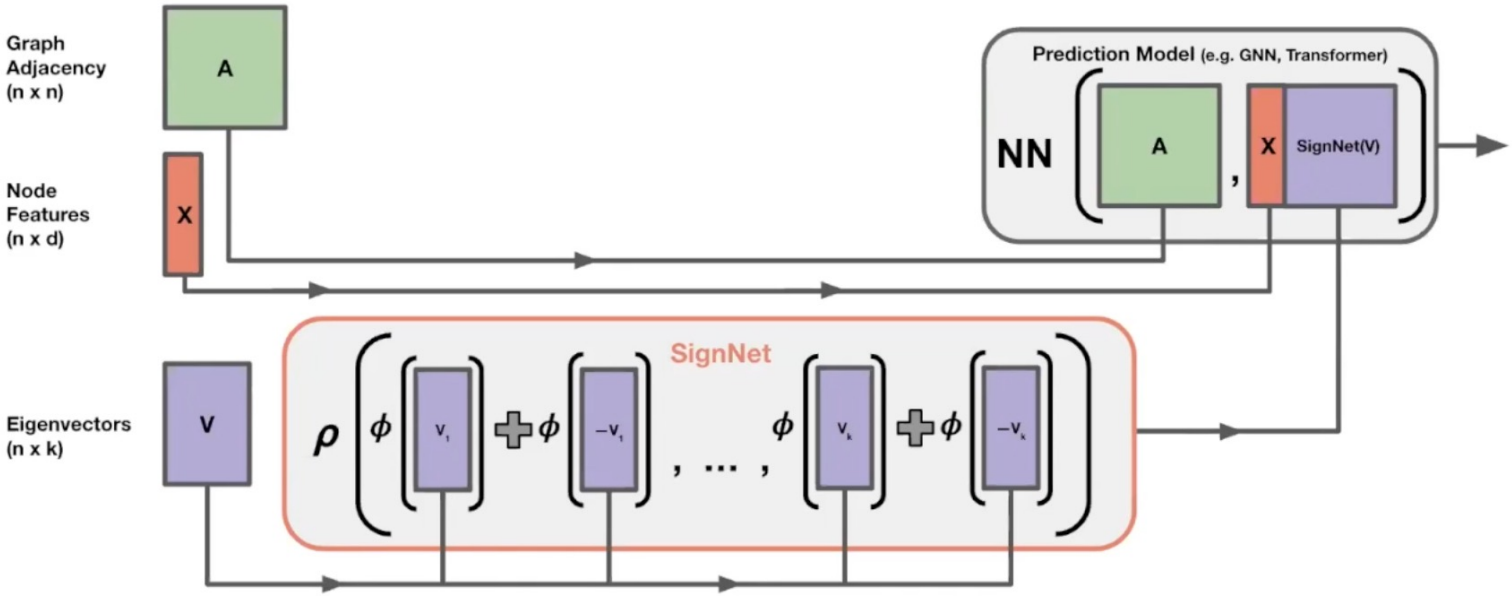
- ρ = MLP, DeepSets, Transformer
- $\emptyset$ = Invariant Graph Network (Maron et al 2018)

Theoretical and Empirical Strength for Graph Learning

Node Positional Encodings



Node Positional Encodings



Experiments: Zinc Dataset

- Beats all eigenvector baselines
- Best positional encoding

Table 1: Results on the ZINC dataset with a 500k parameter budget. All models use edge features. Numbers are the mean and standard deviation over 4 runs, each with different seeds.

Base model	Positional encoding	k	#param	Test MAE (↓)
GatedGCN	No PE	N/A	492k	$0.252_{\pm 0.007}$
	LapPE (flip)	8	492k	$0.198_{\pm 0.011}$
	LapPE (abs.)	8	492k	$0.204_{\pm 0.009}$
	LapPE (can.)	8	505k	$0.298_{\pm 0.019}$
	SignNet ($\phi(v)$ only)	8	495k	$0.148_{\pm 0.007}$
	SignNet	8	495k	$0.121_{\pm 0.005}$
	SignNet	All	491k	$\mathbf{0.100}_{\pm \mathbf{0.007}}$
Sparse Transformer	No PE	N/A	473k	$0.283_{\pm 0.030}$
	LapPE (flip)	16	487k	$0.223_{\pm 0.007}$
	SignNet	16	479k	$0.115_{\pm 0.008}$
	SignNet	All	486k	$\mathbf{0.102}_{\pm \mathbf{0.005}}$
GINE	No PE	N/A	470k	$0.170_{\pm 0.002}$
	LapPE (flip)	16	470k	$0.178_{\pm 0.004}$
	SignNet	16	470k	$0.147_{\pm 0.005}$
	SignNet	All	417k	$\mathbf{0.102}_{\pm \mathbf{0.002}}$
PNA	No PE	N/A	474k	$0.133_{\pm 0.011}$
	LapPE (flip)	8	474k	$0.132_{\pm 0.010}$
	SignNet	8	476k	$0.105_{\pm 0.007}$
	SignNet	All	487k	$\mathbf{0.084}_{\pm \mathbf{0.006}}$

SignNet Visualization on Molecules

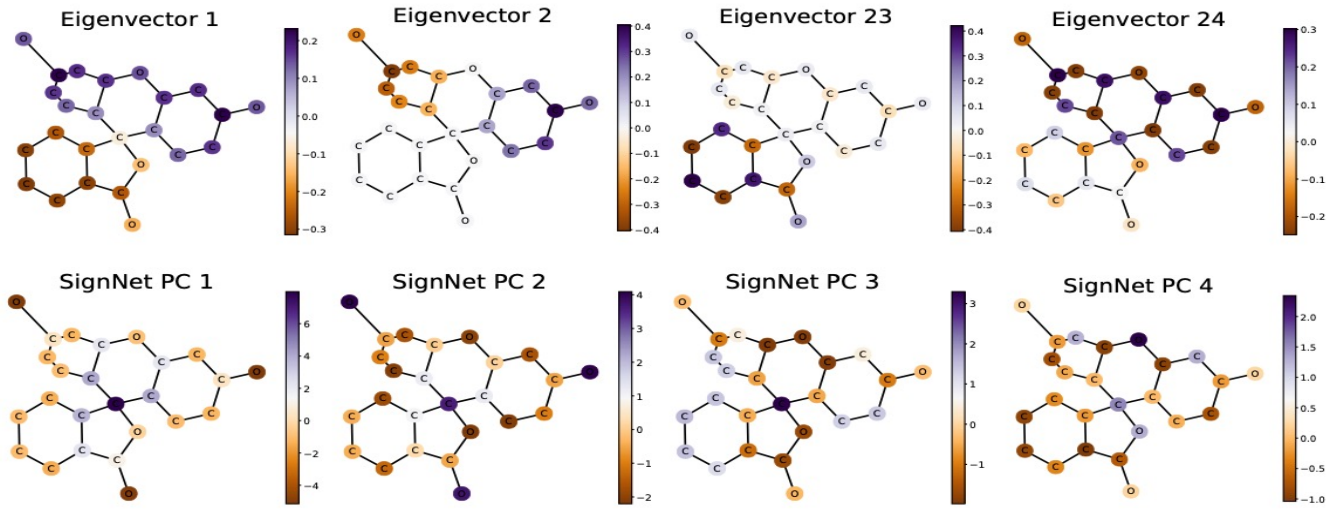


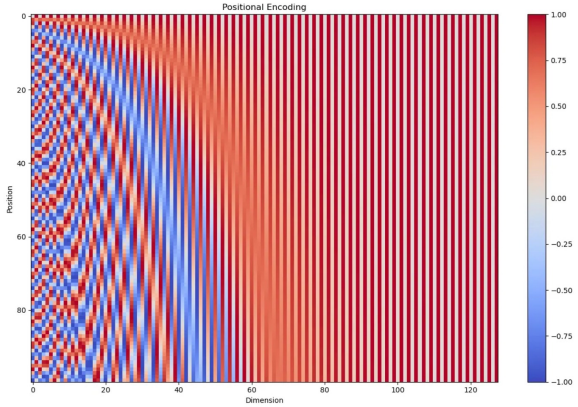
Figure 9: Normalized Laplacian eigenvectors and learned positional encodings for the graph of fluorescein. (Top row) From left to right: smallest and second smallest nontrivial eigenvectors, then second largest and largest eigenvectors. (Bottom row) From left to right: first four principal components of the output $\rho([\phi(v_i) + \phi(-v_i)]_{i=1,\dots,n})$ of SignNet.

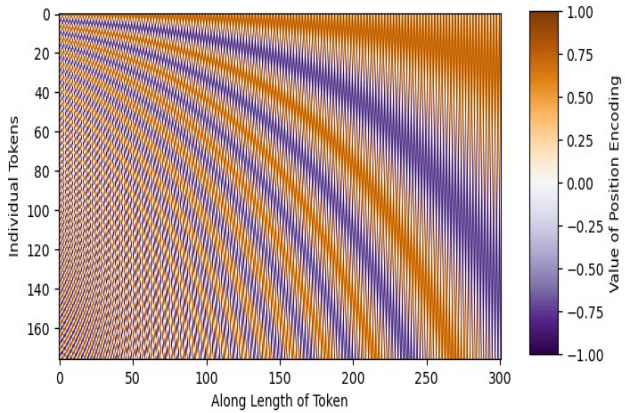
SignNet trained on Zinc dataset learns structural information not found in even original Laplacian eigenvectors!

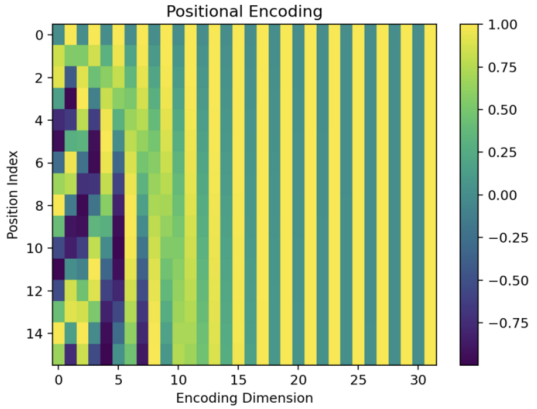
Expressive Power: Existing Positional Encodings

Ability to universally approximate graph encodings like:

- $Heat\ Kernel\ PE_{(Feldman\ et\ al,2022)}$ — SignNet, BasisNet
- $Random\ Walk\ PE_{(Feldman\ et\ al,2022)}$ — SignNet, BasisNet
- $Diffusion, p-step\ random\ walk\ PE_{(Mialon\ et\ al,2022)}$ — BasisNet
- $Generalized\ PageRank, landing\ probability\ distance\ encodings_{(Li\ et\ al,2019)}$ — BasisNet







ETH zürich

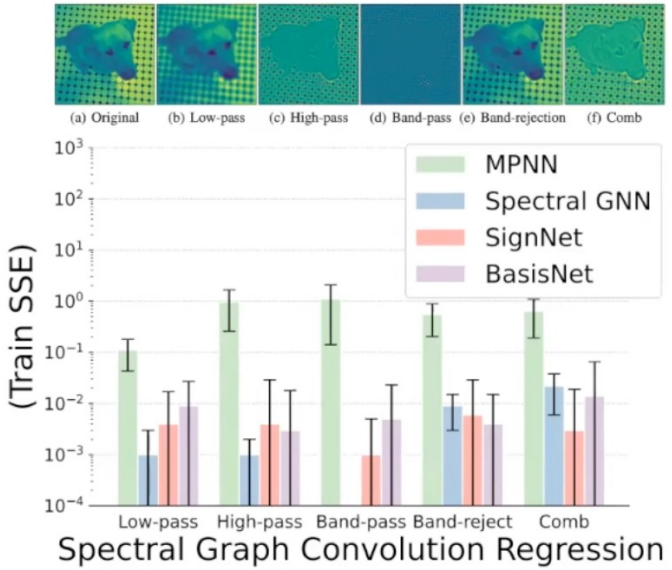
Expressive Power: Spectral Graph Convolutions

SignNet universally approximates spectral graph convolutions

$$\mathrm{h}(\mathrm{V}, \Lambda, X) = \mathrm{V}_p(\Lambda)V^T X = \Sigma_{i=1}^{n} p(\lambda_i) v_i v_i^T X$$

SignNet with: $\emptyset(v, \lambda, X) = \frac{1}{2} p(\lambda) v v^T X, \ \rho = \Sigma_{i=1}^{n}$

- Beat sign flip, absolute value eigvec baselines
- Beat message passing, often beat spectral GNNs



Strengths and Weaknesses

- Strengths
  - Mathematical ideas are decomposed cleanly – first for one dimension and then for multiple dimensions
  - Quantifiable real-world impact across wide range of Positional Encodings and by extension, across wide range of models as well
  - The core idea itself is generalizable across domains – which signals a natural fit for the problem
- Weaknesses
  - The evaluation is mostly focused on a single dataset (the Zinc dataset)
  - The parameter budget could be varied to see the effect of model complexity and how the solution scales with it
  - A more general limit – high computational complexity $\left(O(n^3)\right)$ for computing eigenvectors via eigen decomposition

Summary and Conclusion

- SignNet and BasisNet: Invariant neural nets for eigenvector data
- Generic Approximation: Decomposition theorem into single subspaces
- Ability for Expressiveness: Spectral Graph Convolutions
- Results: Generic and useful positional encoding, State of The Art on Zinc 500k param

Questions?