CS224W Homework 3

Due: November 14, 2024

1 GNNs as MLP of eigenvectors [20 points]

1.1 Batch Node Update [2 points]

Consider the update for Graph Isomorphism Network:

$$\mathbf{x}_v^{(l+1)} = \texttt{MLP}\left((1+\epsilon)\,\mathbf{x}_v^{(l)} + \sum_{u \in \mathcal{N}(v)} \mathbf{x}_u^{(l)}\right), \tag{1}$$

where $\mathbf{x}_v^{(l)} \in \mathbb{R}^{d_l}$ is the embedding of node v at layer l. Let $\mathbf{X}^{(l)} \in \mathbb{R}^{N \times d_l}$ be a matrix containing the embeddings of all the nodes in the graph, i.e., $\mathbf{X}^{(l)}[:, v] = \mathbf{x}_v^{(l)}$. Also, let $\mathbf{A} \in \{0,1\}^{N \times N}$ represent the adjacency matrix of the graph. Write down the update of $\mathbf{X}^{(l+1)}$ as a function of $\mathbf{X}^{(l)}$ and $\mathbf{A}$.

★ **Solution** ★

$$\mathbf{X}^{(l+1)} = \texttt{MLP}\left(\left((1+\epsilon)\,\mathbf{I} + \mathbf{A}\right)\mathbf{X}^{(l)}\right)$$

1.2 Single Layer MLP [2 points]

Assume that $\texttt{MLP}\,()$ represents a single layer MLP with no bias term. Write down the update of $\mathbf{X}^{(l+1)}$ as a function of $\mathbf{X}^{(l)}$ and $\mathbf{A}$, and the trainable parameters $\mathbf{W}^{(l)}$ of layer l.

★ **Solution** ★

$$\mathbf{X}^{(l+1)} = \sigma\left(\left(\mathbf{A} + (1+\epsilon)\;\mathbf{I}\right)\;\mathbf{X}^{(l)}\mathbf{W}^{(l)}\right)$$

1.3 Eigenvector Extension [4 points]

Let $\{\lambda_n, \mathbf{v}_n\}_{n=1}^{N}$ represent the eigenvalues and eigenvectors of the graph adjacency. Then we can write $\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$, where $\mathbf{V} \in \mathbb{R}^{N\times N}$ is the matrix of eigenvectors with $\mathbf{V}[:, n] = \mathbf{v}_n$ and $\boldsymbol{\Lambda} \in \mathbb{R}^{N\times N}$ is the diagonal matrix of eigenvalues with $\boldsymbol{\Lambda}[n, n] = \lambda_n$. Show that

$$\mathbf{X}^{(l+1)} = \sigma\left(\mathbf{V}\hat{\mathbf{W}}^{(l)}\right), \quad \hat{\mathbf{W}}^{(l)}[n, j] = (\lambda_n + 1 + \epsilon)\sum_{i=1}^{d_l} \mathbf{W}^{(l)}[i, j]\langle \mathbf{v}_n, \mathbf{X}^{(l)}[:, i]\rangle,$$

where $\langle\cdot\rangle$ denotes the dot product. Hint: Use the fact that the eigenvectors are orthonormal. Next, show that each feature across all nodes, $\mathbf{X}^{(l+1)}[:, i]$, can be expressed as a linear combination of eigenvectors, followed by a pointwise nonlinearity.

★ **Solution** ★

$$\begin{aligned}
\hat{\mathbf{W}}^{(l)}[n, j] &= (\lambda_n + 1 + \epsilon)\sum_{i=1}^{d_l} \mathbf{W}^{(l)}[i, j]\langle \mathbf{v}_n, \mathbf{X}^{(l)}[:, i]\rangle \\
&= (\lambda_n + 1 + \epsilon)\sum_{i=1}^{d_l} \langle \mathbf{v}_n, \mathbf{X}^{(l)}[:, i]\rangle \mathbf{W}^{(l)}[i, j] \\
&= (\lambda_n + 1 + \epsilon)\sum_{i=1}^{d_l} \left(\mathbf{v}_n^{\mathbf{T}}\, \mathbf{X}^{(l)}[:, i]\right) \mathbf{W}^{(l)}[i, j] \\
&= (\lambda_n + 1 + \epsilon)\ \mathbf{v}_n^{\mathbf{T}}\, \mathbf{X}^{(l)}\, \mathbf{W}^{(l)}[:, j] \\
\hat{\mathbf{W}}^{(l)}[n, :] &= (\lambda_n + 1 + \epsilon)\ \mathbf{v}_n^{\mathbf{T}}\, \mathbf{X}^{(l)}\, \mathbf{W}^{(l)} \\
\hat{\mathbf{W}}^{(l)} &= (\Lambda + (1 + \epsilon))\ \mathbf{V}^{\mathbf{T}}\, \mathbf{X}^{(l)}\, \mathbf{W}^{(l)} \\
\mathbf{V}\hat{\mathbf{W}}^{(l)} &= \mathbf{V}\,(\Lambda + (1 + \epsilon))\ \mathbf{V}^{\mathbf{T}}\, \mathbf{X}^{(l)}\, \mathbf{W}^{(l)} \\
&= (\mathbf{V}\Lambda + \mathbf{V}\,(1 + \epsilon))\ \mathbf{V}^{\mathbf{T}}\, \mathbf{X}^{(l)}\, \mathbf{W}^{(l)} \\
&= \left(\mathbf{V}\Lambda\mathbf{V}^{\mathbf{T}} + \mathbf{V}\mathbf{V}^{\mathbf{T}}\,(1 + \epsilon)\right)\ \mathbf{X}^{(l)}\, \mathbf{W}^{(l)} \\
&= (\mathbf{A} + (1 + \epsilon)\,\mathbf{I})\ \mathbf{X}^{(l)}\, \mathbf{W}^{(l)} \\
\sigma(\mathbf{V}\hat{\mathbf{W}}) &= \sigma\left((\mathbf{A} + (1 + \epsilon)\,\mathbf{I})\ \mathbf{X}^{(l)}\, \mathbf{W}^{(l)}\right) = \mathbf{X}^{(l+1)}
\end{aligned}$$

1.4 GraphSAGE [4 points]

Perform the same analysis for the GraphSAGE update when the aggregation function is sum pooling. Recall that the GraphSAGE update function is

$$\begin{aligned}\mathbf{x}_v^{(l+1)} &= \sigma\left(\mathbf{W}^{(l)} \cdot \mathrm{CONCAT}\left(\mathbf{x}_v^{(l)}, \mathbf{x}_{N(v)}^{(l)}\right)\right) \\ &= \sigma\left(\mathbf{W}_1^{(l)}\mathbf{x}_v^{(l)} + \mathbf{W}_2^{(l)}\mathrm{AGG}\left(\mathbf{x}_u^{(l)}, \forall u \in N(v)\right)\right)\end{aligned}$$

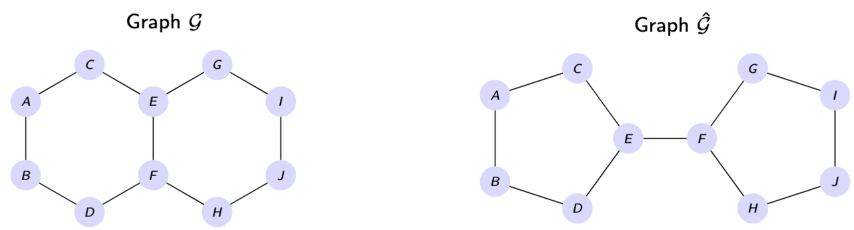
★ **Solution** ★

$$\mathbf{X^{(l+1)}} = \sigma\left(\mathbf{X^{(l)}W_1^{(l)}} + \mathbf{AX^{(l)}W_2^{(l)}}\right) = \sigma\left(\mathbf{V\hat{W}}\right)$$

Where

$$\hat{\mathbf{W}}^{(l)}[n, j] = \sum_{i=1}^{d_l}(\mathbf{W_1^{(l)}}[i, j] + \lambda_n \mathbf{W_2^{(l)}}[i, j])\langle \mathbf{v}_n, \mathbf{X}^{(l)}[:, i]\rangle$$

1.5 Eigendecomposition Analysis [8 points]



For graphs $\mathcal{G}$ and $\hat{\mathcal{G}}$ instantiate the graph adjacencies in Numpy, PyTorch, or PyG, and compute their eigenvalue decompositions. What do you observe?

★ **Solution** ★

For graph $\mathcal{G}$:

Eigenvalue vector $\vec{v}_{\mathcal{G}}$: $\begin{pmatrix} -2.30 \\ 2.30 \\ -1.62 \\ -1.30 \\ -1.00 \\ -0.62 \\ 1.62 \\ 0.62 \\ 1.00 \\ 1.30 \end{pmatrix}$

$$\text{Eigenvector matrix } \mathbf{V}_{\mathcal{G}}: \begin{bmatrix} -0.23 & 0.23 & 0.43 & 0.17 & 0.41 & -0.26 & 0.43 & 0.26 & 0.41 & -0.17 \\ 0.23 & 0.23 & -0.43 & 0.17 & -0.41 & -0.26 & 0.43 & -0.26 & 0.41 & 0.17 \\ 0.30 & 0.30 & -0.26 & -0.40 & -0.00 & 0.43 & 0.26 & 0.43 & 0.00 & -0.40 \\ -0.30 & 0.30 & 0.26 & -0.40 & -0.00 & 0.43 & 0.26 & -0.43 & -0.00 & 0.40 \\ -0.46 & 0.46 & 0.00 & 0.35 & -0.41 & -0.00 & 0.00 & -0.00 & -0.41 & -0.35 \\ 0.46 & 0.46 & -0.00 & 0.35 & 0.41 & 0.00 & -0.00 & -0.00 & -0.41 & 0.35 \\ 0.30 & 0.30 & 0.26 & -0.40 & -0.00 & -0.43 & -0.26 & -0.43 & 0.00 & -0.40 \\ -0.30 & 0.30 & -0.26 & -0.40 & -0.00 & -0.43 & -0.26 & 0.43 & -0.00 & 0.40 \\ -0.23 & 0.23 & -0.43 & 0.17 & 0.41 & 0.26 & -0.43 & -0.26 & 0.41 & -0.17 \\ 0.23 & 0.23 & 0.43 & 0.17 & -0.41 & 0.26 & -0.43 & 0.26 & 0.41 & 0.17 \end{bmatrix}$$

For graph $\hat{\mathcal{G}}$:

$$\text{Eigenvalue vector } \vec{v}_{\hat{\mathcal{G}}}: \begin{pmatrix} 2.30 \\ 1.86 \\ -2.11 \\ -1.30 \\ 1.00 \\ 0.25 \\ -1.62 \\ -1.62 \\ 0.62 \\ 0.62 \end{pmatrix}$$

$$\text{Eigenvalue matrix } \mathbf{V}_{\hat{\mathcal{G}}}: \begin{bmatrix} -0.23 & -0.36 & 0.10 & 0.17 & -0.41 & 0.33 & -0.60 & -0.06 & 0.37 & 0.08 \\ -0.23 & -0.36 & 0.10 & 0.17 & -0.41 & 0.33 & 0.60 & 0.06 & -0.37 & -0.08 \\ -0.30 & -0.31 & -0.30 & -0.40 & -0.00 & -0.25 & 0.37 & 0.03 & 0.60 & 0.14 \\ -0.30 & -0.31 & -0.30 & -0.40 & 0.00 & -0.25 & -0.37 & -0.03 & -0.60 & -0.14 \\ -0.46 & -0.22 & 0.54 & 0.35 & 0.41 & -0.40 & 0.00 & -0.00 & 0.00 & -0.00 \\ -0.46 & 0.22 & -0.54 & 0.35 & 0.41 & 0.40 & -0.00 & -0.00 & 0.00 & -0.00 \\ -0.30 & 0.31 & 0.30 & -0.40 & -0.00 & 0.25 & 0.00 & 0.37 & -0.00 & 0.59 \\ -0.30 & 0.31 & 0.30 & -0.40 & -0.00 & 0.25 & 0.00 & -0.37 & 0.00 & -0.59 \\ -0.23 & 0.36 & -0.10 & 0.17 & -0.41 & -0.33 & -0.00 & -0.60 & -0.00 & 0.36 \\ -0.23 & 0.36 & -0.10 & 0.17 & -0.41 & -0.33 & -0.00 & 0.60 & 0.00 & -0.36 \end{bmatrix}$$

Consider a GIN where all nodes start with the same initial color, i.e., $\mathbf{x}_v^{(0)} = 1$ for all nodes $v \in \mathcal{V}$. This setup is equivalent to having $\mathbf{X}^{(0)} = \mathbf{1}$, where $\mathbf{1}$ denotes the all-one vector. This is the initialization of the WL test. Using the equations in 1.3, derive the expression for $\mathbf{X}^{(1)}$.

★ Solution ★

$$\mathbf{X}^{(1)} = \sigma\left(\mathbf{V}\hat{\mathbf{W}}^{(0)}\right), \quad \hat{\mathbf{W}}^{(0)}[n, j] = (\lambda_n + 1 + \epsilon) \sum_{i=1}^{d_l} \mathbf{W}^{(l)}[i, j] \langle \mathbf{v}_n, \mathbf{X}^{(0)} \rangle,$$

Observe that each column $\mathbf{X}^{(1)}[:, j]$ is a linear combination of eigenvectors, followed by a pointwise nonlinearity. What is the weight associated with each eigenvector? What factors determine this weight?

★ **Solution** ★

The weight associated with the n-th eigenvector in the column j of the new $\mathbf{X}$ matrix is equal to:

$$\hat{\mathbf{W}}^{(0)}[n, j] = (\lambda_n + 1 + \epsilon) \sum_{i=1}^{d_l} \mathbf{W}^{(l)}[i, j] \langle \mathbf{v}_n, \mathbf{X}^{(0)} \rangle,$$

It is determined by the associated eigenvalue, the ϵ parameter, the weight associated with column j in the original layer, and the dot product $\langle \mathbf{v}_n, \mathbf{X}^{(0)} \rangle$.

Compute the dot product $\langle \mathbf{v}_n, \mathbf{X}^{(0)} \rangle$, for each eigenvector across both graphs. What do you observe?

★ **Solution** ★

For graph $\mathcal{G}$: $\begin{pmatrix} 0.00 \\ 3.05 \\ 0.00 \\ -0.21 \\ -0.00 \\ 0.00 \\ 0.00 \\ -0.00 \\ 0.82 \\ -0.00 \end{pmatrix}$

For graph $\hat{\mathcal{G}}$: $\begin{pmatrix} -3.05 \\ -0.00 \\ -0.00 \\ -0.21 \\ -0.82 \\ -0.00 \\ 0.00 \\ 0.00 \\ -0.00 \\ 0.00 \end{pmatrix}$

What does the previous result suggest about $\mathbf{X}^{(1)}$ for the graphs $\mathcal{G}$ and $\hat{\mathcal{G}}$?

★ **Solution** ★

For each eigenvector, the resulting number in the previous calculation is either equal to 0, or equal to a number y such that it is paired with an eigenvector λ_y and the existing pairs $(|y|, \lambda_y)$ (we take the module of y because inverting the eigenvectors doesn't change the corresponding eigenvalues and still keeps the eigenbasis orthonormal) are the same in both graphs. We can check that the eigenvectors corresponding to these eigenvalues are the same across both graphs, so the same eigenvectors will receive the same weights in the MLP layer and thus the resulting $\mathbf{X}^{(1)}$ embeddings will be the same.