Sheaf Neural Networks

THE OHIO STATE UNIVERSITY

Jakob Hansen and Thomas Gebhart

UNIVERSITY OF MINNESOTA

Motivation

The structure of graph neural networks constrain their ability to learn over graphs where relations between nodes are non-constant, asymmetric, and varying in dimension. We present a generalization of graph convolutional networks by generalizing the diffusion operation underlying this class of graph neural networks. These sheaf neural networks are based on the sheaf Laplacian a generalization of the graph Laplacian that encodes additional relational structure parameterized by the underlying graph.

Cellular Sheaves

Sheaves have a long history as central objects of study in algebraic topology and category theory. Cellular sheaves—defined over cell complexes—attach data to the vertices and edges of graphs, *characterizing* relationships represented by the graph. Given an undirected graph G, a cellular sheaf is defined as:

- a vector space $\mathcal{F}(v)$ for each vertex v of G,
- a vector space $\mathcal{F}(e)$ for each edge e of G, and
- a linear map $\mathcal{F}_{v \trianglelefteq e} : \mathcal{F}(v) \to \mathcal{F}(e)$ for each incident vertex-edge pair $v \trianglelefteq e$

Sheaf Neural Networks

Cellular sheaves inherit a coboundary operator δ, mapping from the data space of all vertices to the data space of all edges:

$$(\delta x)_e = \mathcal{F}_{v \trianglelefteq e} x_v - \mathcal{F}_{u \trianglelefteq e} x_u \text{ where } e = u \to v$$

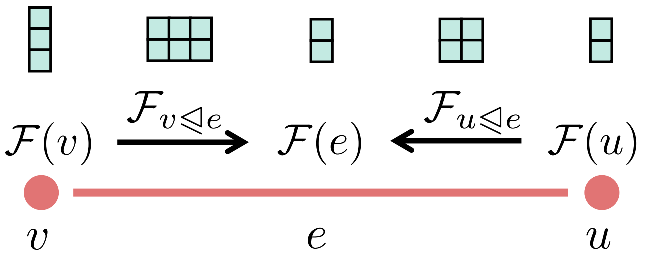
The *sheaf Laplacian* $L_{\mathcal{F}} = \delta^T \delta$ is PSD with kernel the subspace of globally-consistent data.

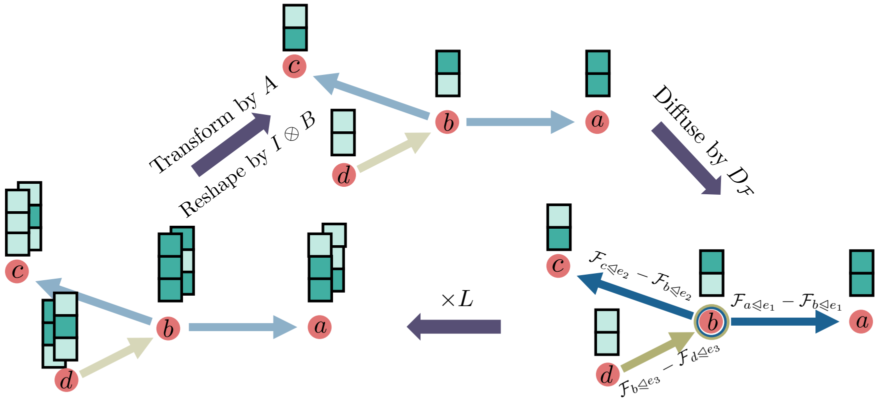
$$L_{\mathcal{F}} = \delta^T \delta = \begin{pmatrix} \mathcal{F}_{v \trianglelefteq e}^T \mathcal{F}_{v \trianglelefteq e} & -\mathcal{F}_{v \trianglelefteq e}^T \mathcal{F}_{u \trianglelefteq e} \\ -\mathcal{F}_{u \trianglelefteq e}^T \mathcal{F}_{v \trianglelefteq e} & \mathcal{F}_{u \trianglelefteq e}^T \mathcal{F}_{u \trianglelefteq e} \end{pmatrix}$$

Assume N_v nodes in the graph each with N_{feat}, k-dimensional features. Concatenating node features into input matrix X of size $N_v k \times N_{\text{feat}}$, we define a sheaf-convolutional layer:

$$\text{SheafConv}(A, B)(X) = \rho\,(D_{\mathcal{F}}(I \otimes B) X A)$$

A of size $(N_{\text{feat}}^{\text{in}} \times N_{\text{feat}}^{\text{out}})$ and B of size $(k \times k)$ are learnable parameters. $D_{\mathcal{F}} = I - \frac{1}{d_{\max}} L_{\mathcal{F}}$ is the sheaf diffusion operator, producing a new set of features for each node as a mixture of the previous features according to the sheaf structure.





Experiments

Sheaf neural networks outperform graph convolutional networks on signed graphs where relations are asymmetric.