p-Laplacian Based Graph Neural Networks

Zepeng Zhang

This paper [1] considered the problem of semi-supervised node classification on heterophilic graphs and graphs with noisy edges. To this end, this paper derived a novel p-Laplacian message passing scheme from a discrete regularization framework and proposed a new ^{p}GNN architecture, which can be theoretically verified as an approximation of a polynomial graph filter defined on the spectral domain of the p-Laplacian. They theoretically demonstrate their method works as low-pass and high-pass filters and thereby applicable to both homophilic and heterophilic graphs.

1 Preliminaries and Background

Definition 1 (Graph Gradient)**.** Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a function $f : \mathcal{V} \to \mathbb{R}$, the graph gradient is an operator $\nabla : \mathcal{F}_{\mathcal{V}} \to \mathcal{F}_{\mathcal{E}}$ defined by

$$(\nabla f)([i,j]) = \sqrt{\frac{W_{i,j}}{D_{j,j}}} f(j) - \sqrt{\frac{W_{i,j}}{D_{i,i}}} f(i), \quad \forall [i,j] \in \mathcal{E}. \tag{1}$$

For $[i,j] \notin \mathcal{E}$, $(\nabla f)([i,j]) = 0$. The graph gradient of a function f at vertex i is defined to be $(\nabla f)(i) = ((\nabla f)([i,1]), \ldots, (\nabla f)([i,N]))$ and its Frobenius norm is given by $\|(\nabla f)(i)\|_2 = \left(\sum_{j=1}^{N} (\nabla f)^2([i,j])\right)^{\frac{1}{2}}$, which measures the variation of f around node i. We measure the variation of f over the whole graph $\mathcal{G}$ by $\mathcal{S}_p(f)$ where it is defined to be

$$\mathcal{S}_p(f) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \|(\nabla f)([i,j])\|^p = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \left\| \sqrt{\frac{W_{i,j}}{D_{j,j}}} f(j) - \sqrt{\frac{W_{i,j}}{D_{i,i}}} f(i) \right\|^p, \quad \text{for } p \geq 1.$$

Definition 2 (Graph Divergence)**.** Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and functions $f : \mathcal{V} \to \mathbb{R}$, $g : \mathcal{E} \to \mathbb{R}$, the graph divergence is computed as

$$(\mathsf{div} g)(i) = \sum_{j=1}^{N} \sqrt{\frac{W_{i,j}}{D_{i,i}}} \left(g([i,j]) - g([j,i])\right). \tag{2}$$

Definition 3 (Graph p-Laplacian Operator)**.** Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a function $f : \mathcal{V} \to \mathbb{R}$, the graph p-Laplacian is an operator $\Delta_p : \mathcal{F}_{\mathcal{V}} \to \mathcal{F}_{\mathcal{V}}$ defined by

$$\Delta_p f = -\frac{1}{2}\mathsf{div}\left(\|\nabla f\|^{p-2} \nabla f\right), \quad \text{for } p \geq 1, \tag{3}$$

where $\|\cdot\|^{p-2}$ is element-wise.

Preprint. Under review.

Substituting Eq. (1) and Eq. (2) into Eq. (3), we obtain

$$
\begin{aligned}
&(\Delta_p f)(i) \\
=& -\frac{1}{2}\mathsf{div}\left(\|\nabla f\|^{p-2}\nabla f\right)(i) \qquad (4)\\
=& -\frac{1}{2}\sum_{j=1}^{N}\sqrt{\frac{W_{i,j}}{D_{i,i}}}\left(\left(\|\nabla f\|^{p-2}\nabla f\right)([i,j]) - \left(\|\nabla f\|^{p-2}\nabla f\right)([j,i])\right) \\
=& \frac{1}{2}\sum_{j=1}^{N}\sqrt{\frac{W_{i,j}}{D_{i,i}}}\left(\|\nabla f([j,i])\|^{p-2}\left(\left(\sqrt{\frac{W_{i,j}}{D_{i,i}}}f(i) - \sqrt{\frac{W_{i,j}}{D_{j,j}}}f(j)\right)\right) - \|\nabla f[i,j]\|^{p-2}\left(\sqrt{\frac{W_{i,j}}{D_{j,j}}}f(j) - \sqrt{\frac{W_{i,j}}{D_{i,i}}}f(i)\right)\right) \\
=& \sum_{j=1}^{N}\sqrt{\frac{W_{i,j}}{D_{i,i}}}\left(\|\nabla f[j,i]\|^{p-2}\left(\sqrt{\frac{W_{i,j}}{D_{i,i}}}f(i) - \sqrt{\frac{W_{i,j}}{D_{j,j}}}f(j)\right)\right). \qquad (5)
\end{aligned}
$$

When $p=2$, $\Delta_2 = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ is referred as the ordinary Laplacian operator. When $p=1$, $\Delta_1 = -\frac{1}{2}\mathsf{div}\left(\|\nabla f\|^{-1}\nabla f\right)$ is referred as the curvature operator. Note that Laplacian Δ_2 is a linear operator, while in general for $p \neq 2$, p-Laplacian is nonlinear since $\Delta_p(af) \neq a\Delta_p(f)$ for $a \in \mathbb{R}$.
The graph p-Laplacian is semi-definite:

$$
\begin{aligned}
\langle f, \Delta_p f\rangle =& \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sqrt{\frac{W_{i,j}}{D_{i,i}}}\|\nabla f([j,i])\|^{p-2}\left(\sqrt{\frac{W_{i,j}}{D_{i,i}}}f(i) - \sqrt{\frac{W_{i,j}}{D_{j,j}}}f(j)\right)f(i) \\
&- \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sqrt{\frac{W_{i,j}}{D_{j,j}}}\|\nabla f[j,i]\|^{p-2}\left(\sqrt{\frac{W_{i,j}}{D_{i,i}}}f(i) - \sqrt{\frac{W_{i,j}}{D_{j,j}}}f(j)\right)f(j) \\
=& \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\nabla f([j,i])\|^{p-2}\left(\sqrt{\frac{W_{i,j}}{D_{i,i}}}\sqrt{\frac{W_{i,j}}{D_{i,i}}}f(i)f(i) + \sqrt{\frac{W_{i,j}}{D_{j,j}}}\sqrt{\frac{W_{i,j}}{D_{j,j}}}f(j)f(j) - 2\sqrt{\frac{W_{i,j}}{D_{i,i}}}\sqrt{\frac{W_{i,j}}{D_{j,j}}}f(i)f(j)\right. \\
=& \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\nabla f([j,i])\|^{p-2}\|\nabla f([j,i])\|^{2} \\
=& \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\nabla f([i,j])\|^{p} \\
=& \mathcal{S}_p(f) \geq 0.
\end{aligned}
$$

We also have

$$
\begin{aligned}
\frac{\partial \mathcal{S}_p(f)}{\partial f}\,|_i =& \frac{\partial\left(\sum_{j=1}^{N}\|\nabla f([i,j])\|^{p} - \|\nabla f([i,i])\|^{p}\right)}{\partial f(i)} \\
=& \frac{\partial\left(\sum_{j=1}^{N}\left\|\sqrt{\frac{W_{i,j}}{D_{j,j}}}f(j) - \sqrt{\frac{W_{i,j}}{D_{i,i}}}f(i)\right\|^{p}\right)}{\partial f(i)} \\
=& \sum_{j=1}^{N}\frac{p}{2}\left\|\sqrt{\frac{W_{i,j}}{D_{j,j}}}f(j) - \sqrt{\frac{W_{i,j}}{D_{i,i}}}f(i)\right\|^{p-2} 2\sqrt{\frac{W_{i,j}}{D_{i,i}}}\left(\sqrt{\frac{W_{i,j}}{D_{i,i}}}f(i) - \sqrt{\frac{W_{i,j}}{D_{j,j}}}f(j)\right) \\
=& p(\Delta_p f)(i).
\end{aligned}
$$

2 p-Laplacian Based Graph Neural Networks

The p-Laplacian regularization problem is defined to be

$$\mathbf{F}^{\star} = \arg\min_{\mathbf{F}} \mathcal{L}_p(\mathbf{F}) := \arg\min_{\mathbf{F}} \mathcal{S}_p(\mathbf{F}) + \mu \sum_{i=1}^{N} \left\| \mathbf{F}_{i,:} - \mathbf{X}_{i,:} \right\|^2 , \tag{6}$$

where $\mu \in (0, \infty)$. Different choices of p result in different smoothness constraint on the signals.

With $p = 2$, the p-Laplacian regularization problem has closed-form solution which corresponds to PPNP.

For $p > 1$, the gradient of $\mathcal{L}_p(\mathbf{F})$ over $\mathbf{F}_{i,:}$ for all $i \in [N]$ is

$$\frac{\partial \mathcal{L}_p(\mathbf{F})}{\partial \mathbf{F}_{i,:}} = p \sum_{j=1}^{N} \frac{W_{i,j}}{\sqrt{D_{i,i}}} \left(\left\| \sqrt{\frac{W_{i,j}}{D_{i,i}}} \mathbf{F}_{i,:} - \sqrt{\frac{W_{i,j}}{D_{j,j}}} \mathbf{F}_{j,:} \right\|^{p-2} \left(\frac{1}{\sqrt{D_{i,i}}} \mathbf{F}_{i,:} - \frac{1}{\sqrt{D_{j,j}}} \mathbf{X}_{i,:} \right) \right) + \frac{2\mu}{p} \left(\mathbf{F}_{i,:} - \mathbf{X}_{i,:} \right).$$

Using gradient descent with stepsize $\frac{\alpha_{i,i}^{(k)}}{p}$ where

$$\alpha_{i,i}^{(k)} = 1 / \left(\sum_{j=1}^{N} \frac{M_{i,j}^{(k)}}{D_{i,i}} + \frac{2\mu}{p} \right), \quad \forall i \in [N],$$

$$M_{i,j}^{(k)} = W_{i,j} \left\| \sqrt{\frac{W_{i,j}}{D_{i,i}}} \mathbf{F}_{i,:}^{(k)} - \sqrt{\frac{W_{i,j}}{D_{j,j}}} \mathbf{F}_{j,:}^{(k)} \right\|^{p-2}, \quad \forall i, j \in [N],$$

and denoting $\beta_{i,i}^{(k)} = \frac{2\mu}{p} \alpha_{i,i}, \ \ \forall i \in [N]$, then we have

$$\begin{aligned}
&\mathbf{F}_{i,:}^{(k+1)} \\
&= \mathbf{F}_{i,:}^{(k)} - \frac{\alpha_{i,i}^{(k)}}{p} \left(p \sum_{j=1}^{N} \frac{W_{i,j}}{\sqrt{D_{i,i}}} \left(\left\| \sqrt{\frac{W_{i,j}}{D_{i,i}}} \mathbf{F}_{i,:}^{(k)} - \sqrt{\frac{W_{i,j}}{D_{j,j}}} \mathbf{F}_{j,:}^{(k)} \right\|^{p-2} \left(\frac{1}{\sqrt{D_{i,i}}} \mathbf{F}_{i,:}^{(k)} - \frac{1}{\sqrt{D_{j,j}}} \mathbf{F}_{j,:}^{(k)} \right) \right) + 2\mu \left(\mathbf{F}_{i,:} - \mathbf{X}_{i,:} \right) \right) \\
&= \mathbf{F}_{i,:}^{(k)} - \alpha_{i,i}^{(k)} \sum_{j=1}^{N} \frac{M_{i,j}^{(k)}}{\sqrt{D_{i,i}}} \left(\frac{1}{\sqrt{D_{i,i}}} \mathbf{F}_{i,:}^{(k)} - \frac{1}{\sqrt{D_{j,j}}} \mathbf{F}_{j,:}^{(k)} \right) - \beta_{i,i}^{(k)} \left(\mathbf{F}_{i,:} - \mathbf{X}_{i,:} \right) \\
&= \left(1 - \alpha_{i,i}^{(k)} \sum_{j=1}^{N} \frac{M_{i,j}^{(k)}}{D_{i,i}} - \beta_{i,i}^{(k)} \right) \mathbf{F}_{i,:}^{(k)} + \alpha_{i,i}^{(k)} \sum_{j=1}^{N} \frac{M_{i,j}^{(k)}}{\sqrt{D_{i,i} D_{j,j}}} \mathbf{F}_{j,:}^{(k)} + \beta_{i,i}^{(k)} \mathbf{X}_{i,:} \\
&= \left(1 - \alpha_{i,i}^{(k)} \left(\sum_{j=1}^{N} \frac{M_{i,j}^{(k)}}{D_{i,i}} + \frac{2\mu}{p} \right) \right) \mathbf{F}_{i,:}^{(k)} + \alpha_{i,i}^{(k)} \sum_{j=1}^{N} \frac{M_{i,j}^{(k)}}{\sqrt{D_{i,i} D_{j,j}}} \mathbf{F}_{j,:}^{(k)} + \beta_{i,i}^{(k)} \mathbf{X}_{i,:} \\
&= \alpha_{i,i}^{(k)} \sum_{j=1}^{N} \frac{M_{i,j}^{(k)}}{\sqrt{D_{i,i} D_{j,j}}} \mathbf{F}_{j,:}^{(k)} + \beta_{i,i}^{(k)} \mathbf{X}_{i,:}.
\end{aligned}$$

Let $\mathbf{M}^{(k)} \in \mathbb{R}^{N \times N}$, $\boldsymbol{\alpha}^{(k)} = \mathsf{diag}\left(\alpha_{1,1}^{(k)}, \ldots, \alpha_{N,N}^{(k)}\right)$, $\boldsymbol{\beta}^{(k)} = \mathsf{diag}\left(\beta_{1,1}^{(k)}, \ldots, \beta_{N,N}^{(k)}\right)$, we can rewrite the above gradient step in matrix form as follows:

$$\mathbf{F}^{(k+1)} = \boldsymbol{\alpha}^{(k)} \mathbf{D}^{-1/2} \mathbf{M}^{(k)} \mathbf{D}^{-1/2} \mathbf{F}^{(k)} + \boldsymbol{\beta}^{(k)} \mathbf{X}.$$

It is easy to see that APPNP is a special case of this message passing scheme with $p = 2$. The convergence can be guaranteed with suitable stepsize.

Using this p-Laplacian message passing scheme, the ^{p}GNN is defined as follows:

$$\begin{aligned}
\mathbf{F}^{(0)} &= \mathsf{ReLU}\left(\mathbf{X} \boldsymbol{\Theta}^{(1)} \right) \\
\mathbf{F}^{(k+1)} &= \boldsymbol{\alpha}^{(k)} \mathbf{D}^{-1/2} \mathbf{M}^{(k)} \mathbf{D}^{-1/2} \mathbf{F}^{(k)} + \boldsymbol{\beta}^{(k)} \mathbf{F}^{(0)}, \quad k = 0, \ldots, K-1, \\
\mathbf{Z} &= \mathsf{softmax}\left(\mathbf{F}^{(K)} \boldsymbol{\Theta}^{(2)} \right).
\end{aligned}$$

Some other theoretical analyses made in the paper such as the upperbound of the risk and the spectral analysis with p-Laplacian are also interesting.

References

[1] Guoji Fu, Peilin Zhao, and Yatao Bian. p-laplacian based graph neural networks. *arXiv preprint arXiv:2111.07337*, 2021. (document)