Optimization-Induced Graph Implicit Nonlinear Diffusion

Qi Chen[1], Yifei Wang[1], Yisen Wang[2,3], Jiansheng Yang[1], Zhouchen Lin[2,3,4]

[1]School of Mathematical Sciences, Peking University, China
[2]Key Lab. of Machine Perception (MoE), School of Artificial Intelligence, Peking University, China
[3]Institute for Artificial Intelligence, Peking University, China
[4]Peng Cheng Laboratory, China.

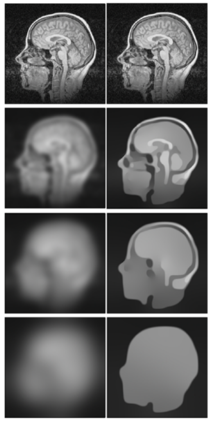
ICML International Conference On Machine Learning

Introduction

Motivation

- Due to the over-smoothing issue, most existing graph neural networks can only capture limited dependencies with their inherently **finite** aggregation layers.
- An implicit layer is equivalent to **infinite** propagation steps, but existing implicit GNNs share a **linear isotropic** aggregation mechanism that treats all neighbors equally, which is the cause of the over-smoothing.

Inspiration from Image Processing



Evolution of an MRI slice under different PDEs (Figure 5.3 & 5.4 in [1])
Left column: Linear diffusion.
Right column: Edge-enhancing anisotropic diffusion.

- Researchers often use the diffusion equation to describe image restoration. If the input image is considered as the initial value, the image will become increasingly blurred along the dimension of time (from top to bottom). We want to preserve desired image structures, such as edges, while blurring others, which is not possible with **linear isotropic** Gaussian blurring (left column).
- Perona-Malik [2] diffusion depends the diffusion coefficient on the norm of the image gradient. They manually design a **nonlinear** function that approximates an impulse function close to edges to reweight the differences between image pixels, so that patterns with larger gradient norms are more easily distinguished from others during the diffusion process (right column).

Contribution

- We develop a new kind of implicit GNNs, GIND, whose **nonlinear** diffusion overcomes the limitations of existing linear isotropic diffusion by adaptively aggregating nonlinear features from neighbors.
- We develop the **first optimization framework** for an **implicit** GNN by showing that the equilibrium states of GIND correspond to the solution of a convex objective. Based on this perspective, we derive principled structural variants with empirical benefits.
- Extensive experiments on node-level and graph-level classification tasks show our GIND obtains state-of-the-art performance among implicit GNNs, and also compares favorably to explicit GNNs.

Method: **G**raph **I**mplicit **N**onlinear **D**iffusion

GIND: Nonlinear Diffusion on Graphs

$$\mathbf{Z} = -\hat{\mathbf{G}}^\top \sigma(\hat{\mathbf{G}}(\mathbf{Z} + b_{\mathbf{\Omega}}(\mathbf{X}))\mathbf{K}^\top)\mathbf{K},$$
$$\hat{\mathbf{Y}} = g_{\mathbf{\Theta}}(\mathbf{X} + \mathbf{Z}),$$

$\boldsymbol{X}$: node feature matrix, $\boldsymbol{Z}$: node equilibrium,
$\boldsymbol{K}$: linear transformation, $\sigma(\cdot)$: element-wise Tanh,
$\widehat{\boldsymbol{G}}$: discrete gradient, $-\widehat{\boldsymbol{G}}^{\boldsymbol{T}}$: discrete divergence,
$\boldsymbol{b}_{\boldsymbol{\Omega}}$: an affine transformation, $\boldsymbol{g}_{\boldsymbol{\Theta}}$: the readout head.

- We design the equilibrium state $\boldsymbol{Z}$ to be the **residual** refinement of the input features $\boldsymbol{X}$ through the diffusion process, *a.k.a.* the transported mass of $\boldsymbol{X}$ [1]. As a result, starting from an initial value, the estimated transported mass $\boldsymbol{Z}$ could be gradually refined through the fixed point iteration of our nonlinear diffusion process. Finally, it will reach a stable equilibrium that cannot be further improved.
- The norm of the flux $\mathrm{j} = -\sigma(G(\cdot)K^T)$ describes the magnitude of the information flow between two nodes. The larger the norm, the higher the degree of mixing between node pairs. We use Tanh to keep small-value gradients while shrinking large-value gradients. In graph diffusion, this amounts to a desirable **anisotropic-like behavior**, which helps prevent over-smoothing and improves robustness to noisy perturbations.

An Optimization Framework for GIND

Theorem 4.1. *Assume that the nonlinear function $\sigma(\cdot)$ is monotone and L_σ-Lipschitz,* i.e.,

$$0 \leq \frac{\sigma(a) - \sigma(b)}{a - b} \leq L_\sigma, \forall\, a, b \in \mathbb{R}, a \neq b, \qquad (11)$$

and $1 \geq L_\sigma \left\|\mathbf{K} \otimes \hat{\mathbf{G}}\right\|_2^2 = L_\sigma \left\|\mathbf{K}\right\|_2^2 \left\|\hat{\mathbf{G}}\right\|_2^2$. *Then there exists a convex function $\varphi(\boldsymbol{z})$, such that its minimizer is the solution to the equilibrium equation $\boldsymbol{z} = f(\boldsymbol{z})$. Furthermore, we have* $\mathrm{Prox}_\varphi(\boldsymbol{z}) = \frac{1}{L_\sigma+1}(L_\sigma \boldsymbol{z} + f(\boldsymbol{z}))$.

The theorem shows that the learned representation can be formalized as the minimizer of an explicit convex optimization objective. With this property, we can induce new structural variants by modifying the corresponding objective. To be specific, we can embed prior properties to the equilibrium, as well as introducing skip connections to promote training stability.

- Optimization-Inspired Skip-Connection: $\boldsymbol{z} = \mathcal{T}(\boldsymbol{z}) := (1-\alpha)\boldsymbol{z} + \alpha f(\boldsymbol{z}),$
- Optimization-Inspired Feature Regularization: $\boldsymbol{z} = \mathcal{T}(\boldsymbol{z}) \circ \mathcal{T}_{\mathcal{R}},$
  - Laplacian Regularization: $\mathcal{R}_{\mathrm{Lap}}(\boldsymbol{z}) = \boldsymbol{z}^\top \tilde{\mathbf{D}}^{-\frac{1}{2}} L \tilde{\mathbf{D}}^{-\frac{1}{2}} \boldsymbol{z} = \left\|\hat{\mathbf{G}}\boldsymbol{z}\right\|^2.$
  - Feature Decorrelation: $\mathcal{R}_{\mathrm{Dec}}(\boldsymbol{z}) = \frac{1}{2}\left\|\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}^\top - \mathbf{I}\right\|_F^2,$ where $\hat{z}$ is the normalized z.

Results

GIND performs well on heterophilic graphs with nonlinear diffusion.

Type	Method	Cornell	Texas	Wisconsin	Chameleon	Squirrel
Explicit	GCN	59.19±3.51	64.05±5.28	61.17±4.71	42.34±2.77	29.0±1.10
	GAT	59.46±6.94	61.62±5.77	60.78±8.27	46.03±2.51	30.51±1.28
	JKNet	58.18±3.87	63.78±6.30	60.98±2.97	44.45±3.17	30.83±1.65
	APPNP	63.78±5.43	64.32±7.03	61.57±3.31	43.85±2.43	30.67±1.06
	Geom-GCN	60.81	67.57	64.12	60.9	38.14
	GCNII	76.75±5.95	73.51±9.95	78.82±5.74	48.59±1.88	32.20±1.06
	H2GCN	82.22±5.67	84.76±5.57	85.88±4.58	60.30±2.31	40.75±1.44
Implicit	IGNN	61.35±4.84	58.37±5.82	53.53±6.49	41.38±2.53	24.99±2.11
	EIGNN	85.13±5.57	84.60±5.41	86.86±5.54	62.92±1.59	46.37±1.39
	GIND (ours)	**85.68±3.83**	**86.22±5.19**	**88.04±3.97**	**66.82±2.37**	**56.71±2.07**

Results on heterophilic node classification datasets: mean accuracy (%) ± standard deviation over different data splits.

GIND obtains significant improvements on graph-level tasks.

Type	Method	MUTAG	PTC	COX2	PROTEINS	NCI1
Explicit	WL	84.1±1.9	58.0±2.5	83.2±0.2	74.7±0.5	**84.5±0.5**
	DCNN	67.0	56.6	-	61.3	62.6
	DGCNN	85.8	58.6	-	75.5	74.4
	GIN	**89.4±5.6**	64.6±7.0	-	76.2±2.8	82.7±1.7
	FDGNN	88.5±3.8	63.4±5.4	83.3±2.9	76.8±2.9	77.8±1.6
Implicit	IGNN*	76.0±13.4	60.5±6.4	79.7±3.4	76.5±3.4	73.5±1.9
	CGS	**89.4±5.6**	64.7±6.4	-	76.3±4.9	77.6±2.0
	GIND (ours)	89.3±7.4	**66.9±6.6**	**84.8±4.2**	**77.2±2.9**	78.8±1.7

Results of graph classification: mean accuracy (%) ± standard deviation over 10 random data splits.

The optimization-induced variants can boost the performance.

Reg	Cora	CiteSeer	PubMed
None	88.25±1.25	76.81±1.68	89.22±0.40
$\mathcal{R}_{\mathrm{Lap}}$	**88.33±1.15**	**76.95±1.73**	**89.42±0.50**
$\mathcal{R}_{\mathrm{Dec}}$	88.29±0.92	76.84±1.70	89.28±0.41

Comparison of different regularization types.

References

[1] Weickert, J. Anisotropic diffusion in image processing. Teubner, 1998.

[2] Perona, P. and Malik, J. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7):629–639, 1990.

[3] Xie, X., Wang, Q., Ling, Z., Li, X., Liu, G., and Lin, Z. Optimization induced equilibrium networks: an explicit optimization perspective for understanding equilibrium models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.