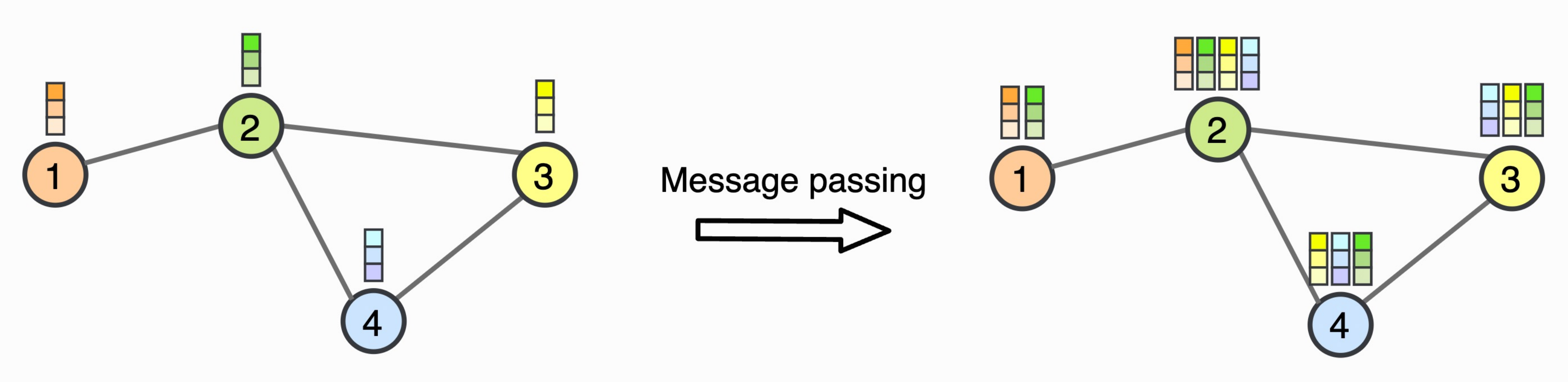
# Optimization-Induced Graph Implicit Nonlinear Diffusion

Qi Chen, Yifei Wang, Yisen Wang, Jiansheng Yang, Zhouchen Lin

# Graph Neural Networks (GNNs)



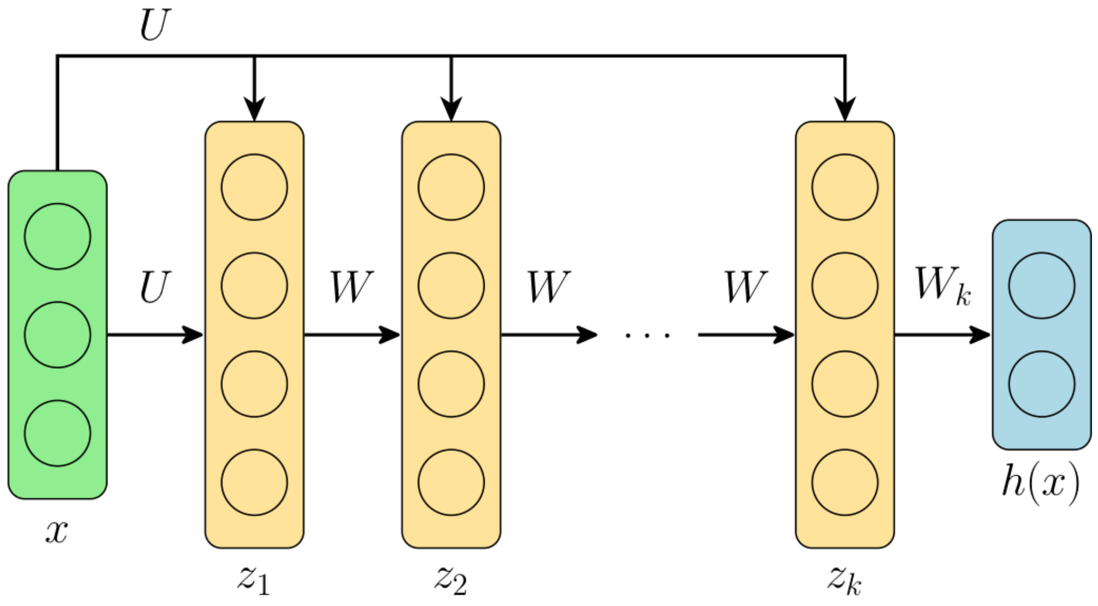

- MPNNs can only aggregate information from $T$-hop neighbors with $T$ propagation steps!
- Problem:
  - $T$ cannot be large due to the **Over-Smoothing** problem.
  - The ability to capture global information is limited by the **FINITE** propagation steps.
- A possible solution:

GNNs with **INFINITE** layers: Implicit GNNs

https://uvadlc-notebooks.readthedocs.io/en/latest/tutorial_notebooks/tutorial7/GNN_overview.html

# Implicit GNNs: GNNs with Infinite Layers

- Implicit Models and Fixed Point Equations
  - An example: IGNN
    - A typical weight-tied k-layer GNN: $Z_k = \sigma(AZ_{k-1}W + UX), k = 1,2,\cdots,n$




What would happen if we were to repeat this update an infinite number of times?

$\Longrightarrow$

A **fixed point equation**

$\boldsymbol{Z = \sigma(AZW + UX)}$

Bai, S., Zico Kolter, J., & Koltun, V. Deep equilibrium models. NeurIPS 2019..

# Implicit GNNs: GNNs with Infinite Layers

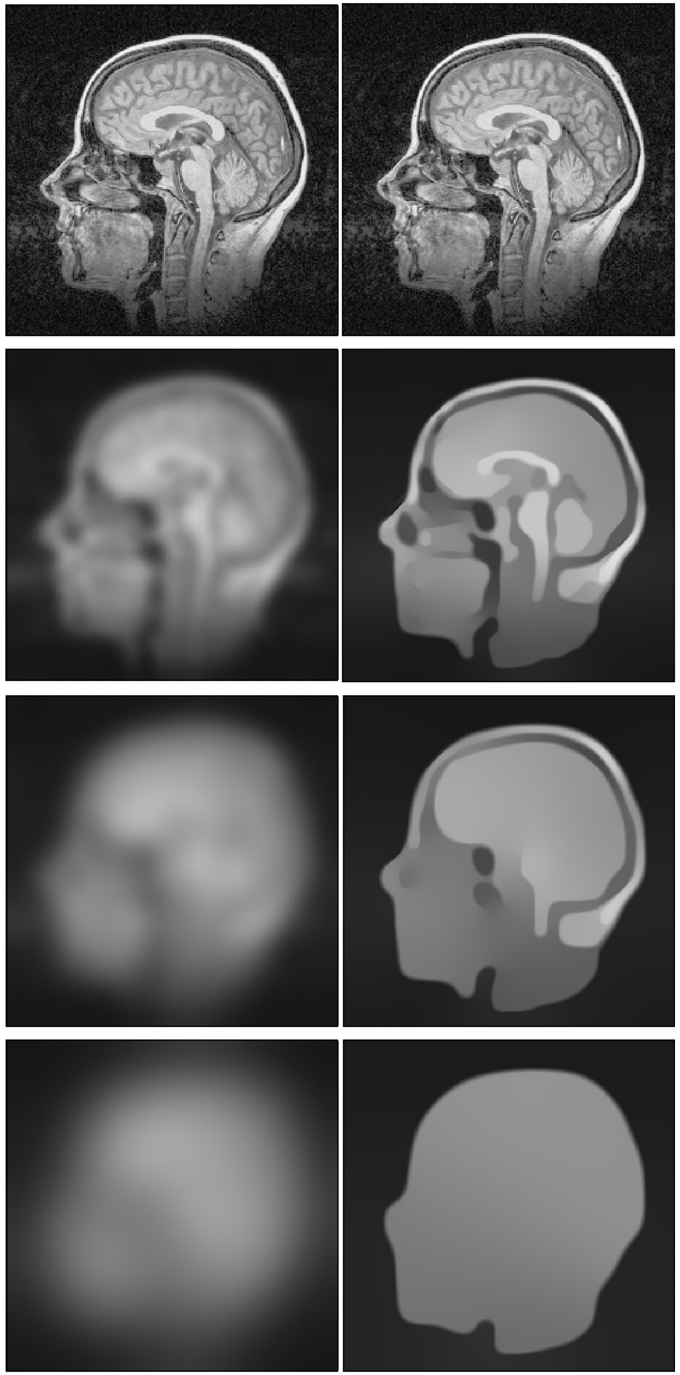
- Implicit Models and Fixed Point Equations
  - An example: IGNN
    - A typical weight-tied k-layer GNN: $Z_k = \sigma(AZ_{k-1}W + UX), k = 1,2,\cdots,n$.
    - The fixed point equation when $\mathrm{n} \to +\infty$:
$$Z = \sigma(AZW + UX).$$
  - Pros
    - **Simple** to construct a model.
    - **Efficient** to calculate the equilibrium.
    - **Global** receptive fields.
  - Cons:
    - Existing implicit GNNs adopt **linear isotropic** diffusion, which is the cause of over-smoothing.

Gu, F., Chang, H., Zhu, W., Sojoudi, S., & El Ghaoui, L. Implicit graph neural networks. NeurIPS 2020.

# Inspirations from PM Diffusion

- Diffusion in image processing
  - Example: Evolution of an MRI slice under different diffusions.
    - Left Column: Linear diffusion.
    - Right Column: Edge-enhancing anisotropic diffusion.
  - Anisotropic diffusion is a cure to over-smoothing.
- Inspiration

  Using **nonlinear** diffusion to construct implicit GNNs!



Weickert, J. (2008). Anisotropic Diffusion in Image Processing.

# Graph Implicit Nonlinear Diffusion (GIND)

- Our fixed point equation

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

$$\hat{\mathbf{Y}} = g_{\mathbf{\Theta}}(\mathbf{X} + \mathbf{Z}),$$

Information exchange between nodes

  - $X$: input feature matrix
  - $Z$: the equilibrium state
  - $Y$: the output feature
  - $\widehat{G}$: normalized incidence matrix (discrete gradient operator)
  - $b_\Omega$: an affine transformation
  - $g_\theta$: the readout head
  - $\sigma$: Tanh

- Anisotropic Property

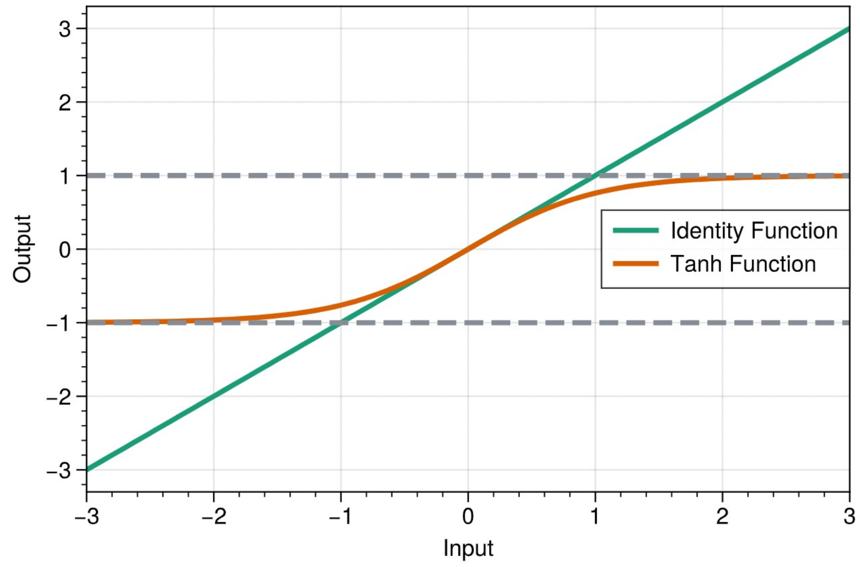



*Figure 1.* Comparison of two activation functions: $\sigma(x) = x$ and $\sigma(x) = \tanh(x)$. The nonlinear activation $\tanh(\cdot)$ keeps small values while shrinking large values.

# Graph Implicit Nonlinear Diffusion (GIND)

- From the optimization perspective
  - The equilibrium of GIND correspond to the solution of a convex objective.

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

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

*and* $1 \geq L_\sigma \left\|\mathbf{K} \otimes \hat{\mathbf{G}}\right\|_2^2 = L_\sigma \|\mathbf{K}\|_2^2 \left\|\hat{\mathbf{G}}\right\|_2^2$. *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* $\mathrm{Prox}_\varphi(\boldsymbol{z}) = \frac{1}{L_\sigma + 1}(L_\sigma \boldsymbol{z} + f(\boldsymbol{z}))$.

# Method: Graph Implicit Nonlinear Diffusion

- From the optimization perspective
  - The equilibrium of GIND correspond to the solution of a convex objective.
  - Optimization-Inspired Variants:
    - Optimization-Inspired Skip-Connection:
      $$z = \Gamma(z) \coloneqq (1-\alpha)z + \alpha f(z)$$
    - Optimization-Inspired Feature Regularization: combining the objective with regularization $\mathcal{R}(z)$ is equivalent to appending one layer before the original layer.
      $$z = \Gamma(z) \circ \Gamma_{\mathcal{R}}, \text{ where } \Gamma_{\mathcal{R}} = Prox_{\mathcal{R}}.$$

# Experiments

- Node-level tasks

*Table 1.* Results on heterophilic node classification datasets: mean accuracy (%) ± standard deviation over different data splits.

| 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** |

# Experiments

- Graph-level tasks

*Table 4.* Results of graph classification: mean accuracy (%) ± standard deviation over 10 random data splits.

| 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 |

# Conclusion

- GIND is the first implicit GNN with nonlinear diffusion.
- GIND has an underlying optimization objective.
- Outperforming SOTA in a variety of tasks.