Digraph Inception Convolutional Networks

Zekun Tong, Yuxuan Liang, Changsheng Sun, Xinke Li, David S. Rosenblum, Andrew Lim

National University of Singapore

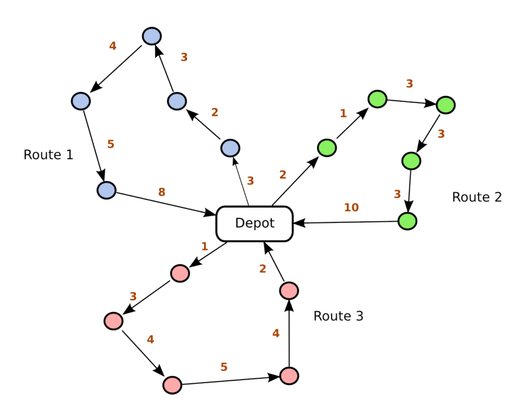
Summary

We make GCNs available in digraphs and propose an Inception network to learn multi-scale features in digraphs.

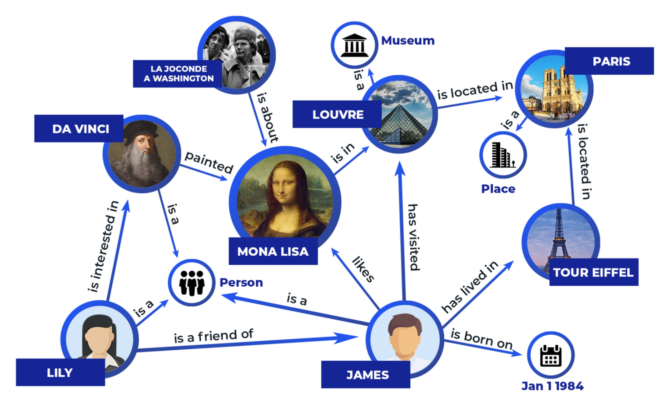
Why digraphs (directed graphs)?

Directed structures are everywhere:

- Recommender Systems
- Traffic Forecasting
- Biology (LGT)
- Neuroscience



Road network

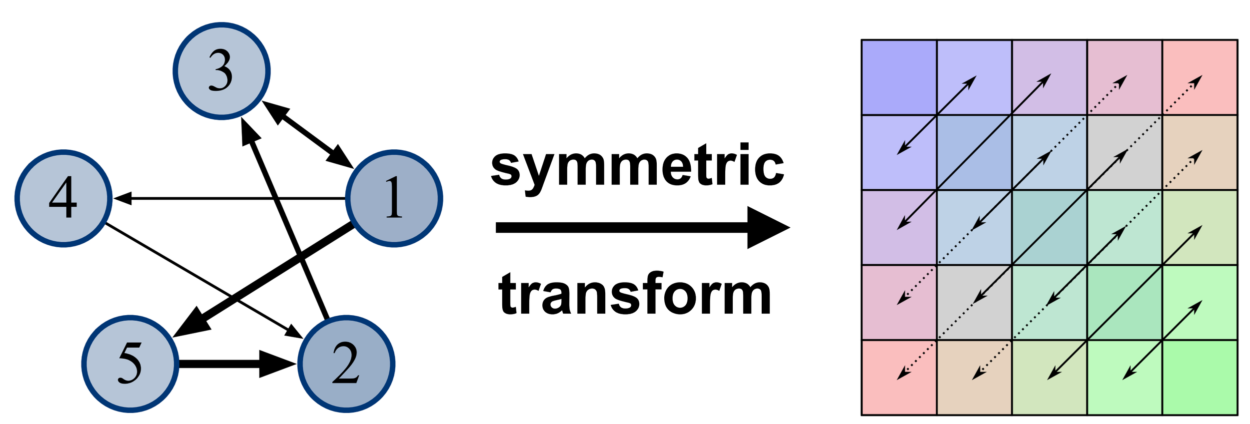


Knowledge graph

Directed graph structure is vital and undeveloped!

Limitations of spectral-based GCNs

Definition Limit

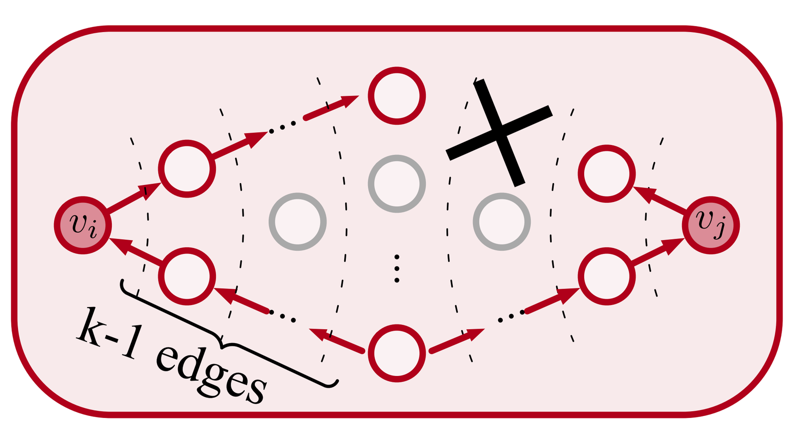


digraph — adjacency matrix

The adjacency matrix needs to be symmetric to use spectral graph convolutions.

Naive symmetric transform is not reasonable.

Structural Limit



long paths only exist between a few points and are often not bidirectional

The unique structure of the digraph makes it difficult to obtain long-range (global) features.

Receptive field would be unbalanced and limited.

Digraph Convolution

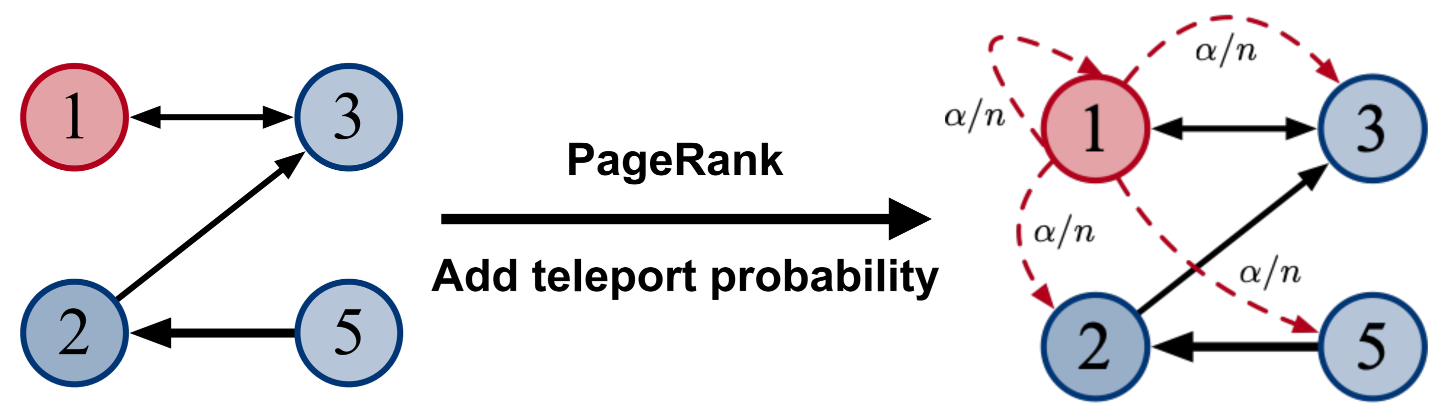
Theoretically define spectral digraph convolution.

Digraph Laplacian based on PageRank:

Given a digraph $\mathcal{G}$ and its adjacency matrix $\mathbf{A}$, $\mathbf{P}_{rw} = \mathbf{D}^{-1}\mathbf{A}$

Using PageRank to make $\mathcal{G}$ irreducible and aperiodic

$$\mathbf{P}_{pr} = (1-\alpha)\mathbf{P}_{rw} + \frac{\alpha}{n}\mathbf{1}^{n\times n} \quad \alpha \text{ is the teleport probability}$$

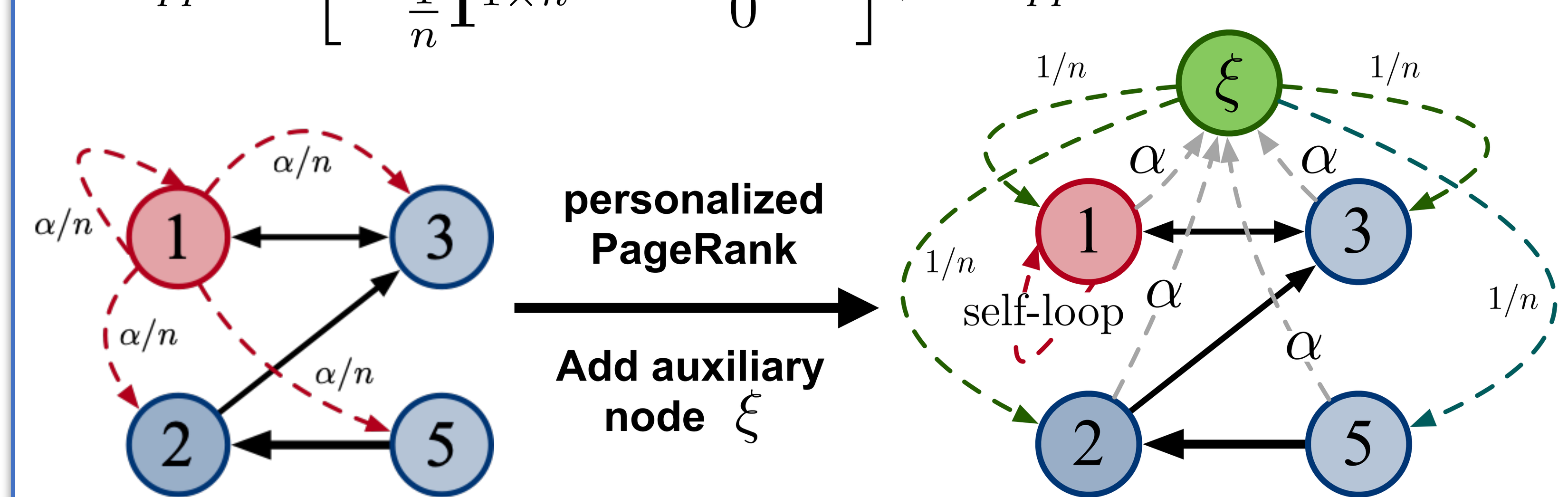


$$\mathcal{L}_{pr} = \mathbf{I} - \frac{1}{2}\left(\mathbf{\Pi}_{pr}^{\frac{1}{2}}\mathbf{P}_{pr}\mathbf{\Pi}_{pr}^{-\frac{1}{2}} + \mathbf{\Pi}_{pr}^{-\frac{1}{2}}\mathbf{P}_{pr}^{T}\mathbf{\Pi}_{pr}^{\frac{1}{2}}\right) \quad \textbf{(Dense)}$$

$\mathbf{\Pi}_{pr}$ is normalized diagonal Perron vector of $\mathbf{P}_{pr}$

Further simplify it using personalized PageRank:

$$\mathbf{P}_{ppr} = \begin{bmatrix} (1-\alpha)\tilde{\mathbf{P}} & \alpha\mathbf{1}^{n\times 1} \\ \frac{1}{n}\mathbf{1}^{1\times n} & 0 \end{bmatrix}, \quad \mathbf{P}_{ppr} \in \mathbb{R}^{(n+1)\times(n+1)}$$



$$\mathcal{L}_{appr} \approx \mathbf{I} - \frac{1}{2}\left(\mathbf{\Pi}_{appr}^{\frac{1}{2}}\tilde{\mathbf{P}}\Pi_{appr}^{-\frac{1}{2}} + \mathbf{\Pi}_{appr}^{-\frac{1}{2}}\tilde{\mathbf{P}}^{T}\mathbf{\Pi}_{appr}^{\frac{1}{2}}\right)$$

Generalize to other forms:

α is the degree of conversion from a directed form to undirected.

$\alpha \to 1$

$$\mathbf{I} - \tilde{\mathbf{D}}^{-1}\tilde{\mathbf{A}} \xleftarrow{\mathcal{G} \text{ is undirected}} \mathbf{I} - \frac{1}{2}\left(\tilde{\mathbf{P}} + \tilde{\mathbf{P}}^{T}\right)$$

undirected random-walk form — trivial-symmetric form

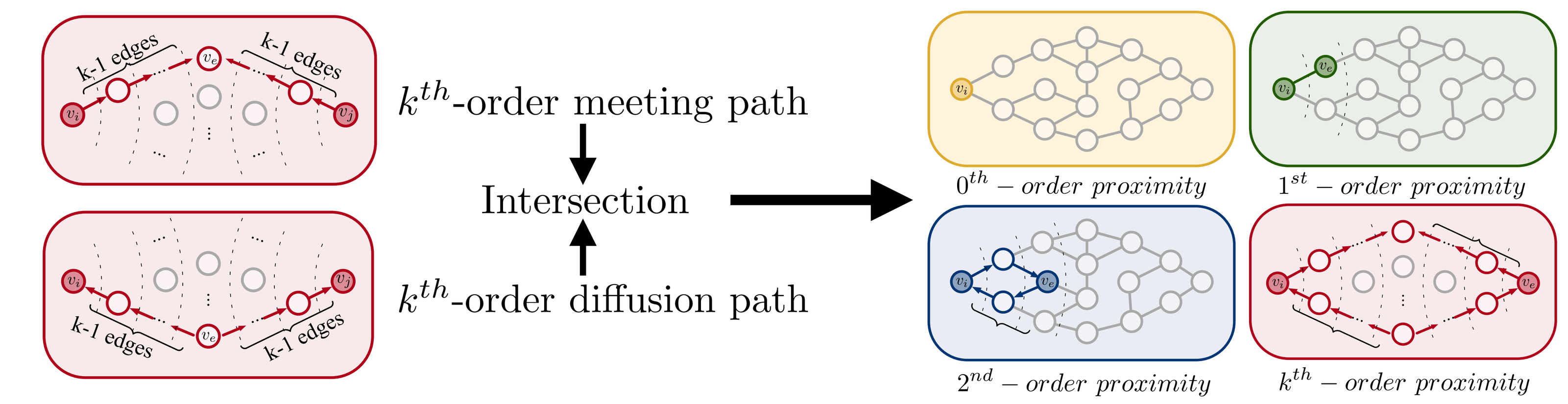
Spectral Digraph Convolution:

$$\mathbf{Z} = \frac{1}{2}\left(\mathbf{\Pi}_{appr}^{\frac{1}{2}}\tilde{\mathbf{P}}\Pi_{appr}^{-\frac{1}{2}} + \mathbf{\Pi}_{appr}^{-\frac{1}{2}}\tilde{\mathbf{P}}^{T}\mathbf{\Pi}_{appr}^{\frac{1}{2}}\right)\mathbf{X}\Theta$$

Digraph Inception Networks

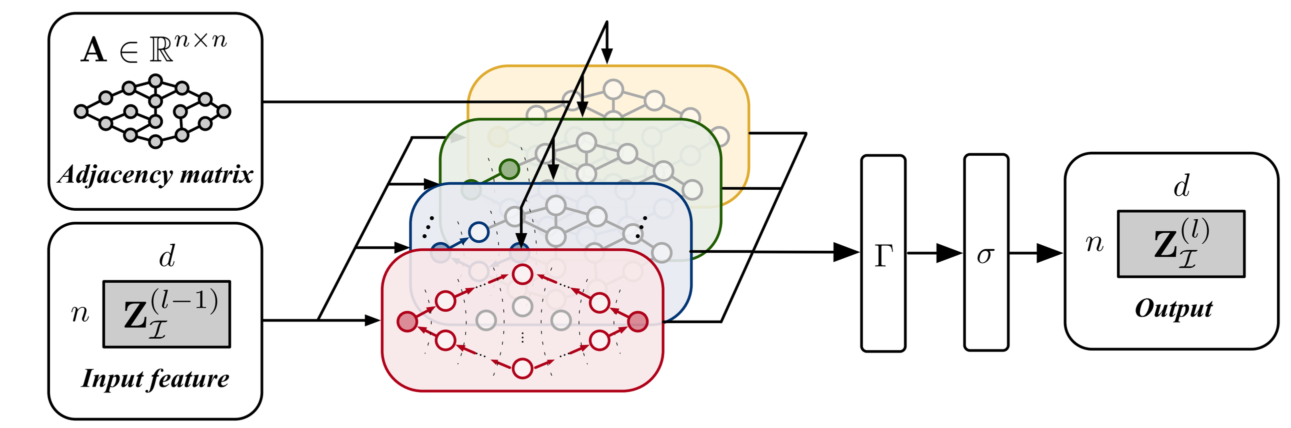
Key idea: friends' friends tend to be mine friends.

Design k^{th}-order Proximity in digraphs



$$\mathbf{P}^{(k)} = \begin{cases} \mathbf{I} & k=0 \\ \tilde{\mathbf{D}}^{-1}\tilde{\mathbf{A}} & k=1 \\ \text{Intersect}\left((\mathbf{P}^{(1)})^{k-1}(\mathbf{P}^{(1)^T})^{k-1}, (\mathbf{P}^{(1)^T})^{k-1}(\mathbf{P}^{(1)})^{k-1}\right)/2 & k \geqslant 2 \end{cases}$$

Multi-scale Conv



Inception block of l^{th} layer

Brief Experimental Results

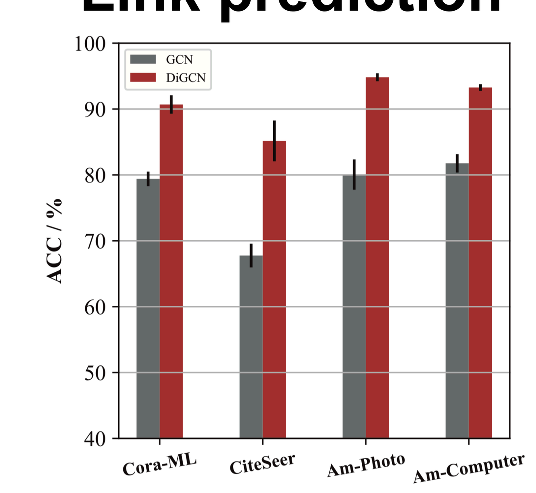
Semi-supervised node classification in digraphs

Table 1: Overall accuracy and training time. "w/ pr" means using $\mathcal{L}_{pr}$; "w/ appr" means using $\mathcal{L}_{appr}$; "w/o IB" means using digraph convolution only; "w/ IB" means using Inception block. The best results are highlighted with **bold** and the second are highlighted with underline.

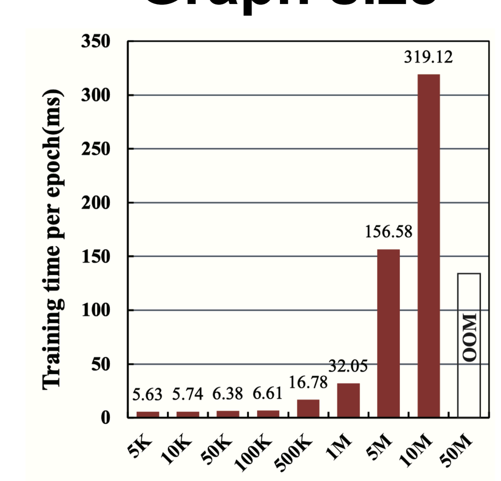
Model	Cora-ML		CiteSeer		Am-Photo		Am-Computer	
	acc	time	acc	time	acc	time	acc	time
ChebNet	64.02 ± 1.5	7.23	56.46 ± 1.4	7.45	80.91 ± 1.0	10.52	73.25 ± 0.8	16.96
GCN	53.11 ± 0.8	4.48	54.36 ± 0.5	4.80	53.20 ± 0.4	4.86	60.50 ± 1.6	5.04
SGC	51.14 ± 0.6	1.92	44.07 ± 3.5	3.58	71.25 ± 1.3	2.31	76.17 ± 0.1	3.68
APPNP	70.07 ± 1.1	6.84	65.39 ± 0.9	6.94	79.37 ± 0.9	6.72	63.16 ± 1.4	6.47
InfoMax	58.00 ± 2.4	4.11	60.51 ± 1.7	4.85	74.40 ± 1.2	31.80	47.32 ± 0.7	41.96
GraphSage	72.06 ± 0.9	6.22	63.19 ± 0.7	6.21	87.57 ± 0.9	8.52	79.29 ± 1.3	14.49
GAT	71.91 ± 0.9	6.02	63.03 ± 0.6	6.12	<u>89.10 ± 0.7</u>	8.83	79.45 ± 1.5	14.66
DGCN	75.02 ± 0.5	6.53	<u>66.00 ± 0.4</u>	6.84	83.66 ± 0.8	36.29	OOM	-
SIGN	66.47 ± 0.9	2.81	60.69 ± 0.4	2.96	74.13 ± 1.0	5.33	69.40 ± 4.8	4.97
Ours								
w/ pr w/o IB	<u>77.11 ± 0.5</u>	39.13	64.77 ± 0.6	47.19	OOM*	-	OOM	-
w/ appr w/o IB	77.01 ± 0.4	2.71	64.92 ± 0.3	2.69	88.72 ± 0.3	2.95	<u>85.55 ± 0.4</u>	4.23
w/ appr w/ IB	**80.28 ± 0.5**	6.38	**66.11 ± 0.7**	6.42	**90.02 ± 0.5**	11.77	**85.94 ± 0.5**	26.63

* OOM stands for out of memory

Link prediction



Graph size



More info