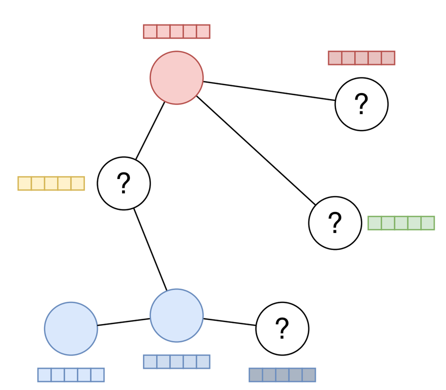
# Analysis of Convolutions, Non-linearity and Depth in Graph Neural Networks using Neural Tangent Kernel

**Mahalakshmi Sabanayagam**, Pascal Esser, Debarghya Ghoshdastidar

## Setup: Node Classification

Graph $G$ with $n$ nodes
Adjacency matrix $A \in \mathbb{R}^{n \times n}$
Degree matrix $D \in \mathbb{R}^{n \times n}$
Feature matrix $X \in \mathbb{R}^{n \times f}$
$m$ node labels $Y \in \{1, \dots, K\}^m$



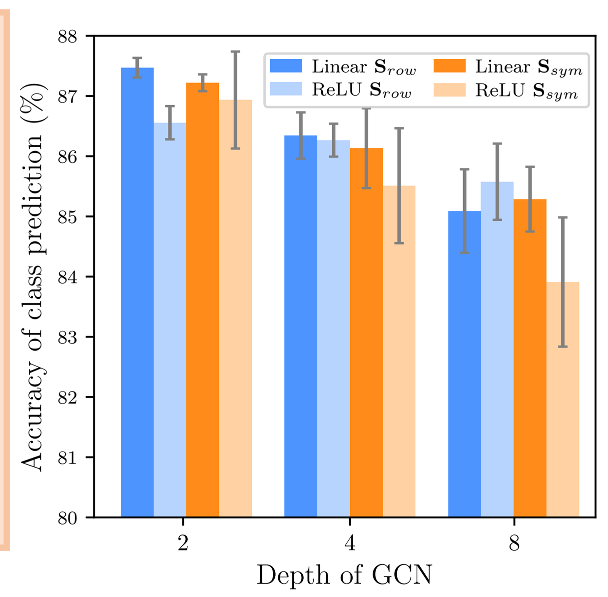
## Intriguing Empirical Observations

### Graph Convolution Network

$$F(S, X) = S\,\sigma\big(\cdots(S\,\sigma(SXW_1)\,W_2)\cdots\big)\,W_d$$

$S = S_{row} = D^{-1}A$ or $S_{sym} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$,
$\sigma(.) =$ Linear or ReLU,
$W_i \in \mathbb{R}^{h \times h}$ are learnable weights

1. $S_{row}$ better than $S_{sym}$
2. Performance ↓ as depth ↑
3. Linear as good as ReLU




**Can we explain the above observations theoretically?**

## Theoretical framework

### Graph Neural Tangent Kernel as $h \to \infty$

$$\Theta^{(d)} = \sum_{k=1}^{d+1} S\Big(\dots S\Big(S\left(\Sigma_k \odot \dot{E}_k\right) S^T \odot \dot{E}_{k+1}\Big) S^T \odot \dots \odot \dot{E}_d\Big) S^T$$

$\Sigma_k$ : Covariance between nodes of layer $k$
$E_k, \dot{E}_k$ : Influence of non-linearity and its derivative

### Degree Corrected Stochastic Block Model

Random graph model characterized by $p, q \in [0,1]$ and $\pi = (\pi_1, \dots, \pi_n) \in [0,1]^n$. Let $r = \dfrac{p-q}{p+q}$.

Then for $K$ latent classes, $C_i \in \{1, \dots, K\}$, the population adjacency matrix $M = \mathbb{E}[A]$ is,

$$M_{ij} = \begin{cases} p\pi_i\pi_j & \text{if } C_i = C_j \\ q\pi_i\pi_j & \text{if } C_i \neq C_j \end{cases}$$

### Our Analysis Framework

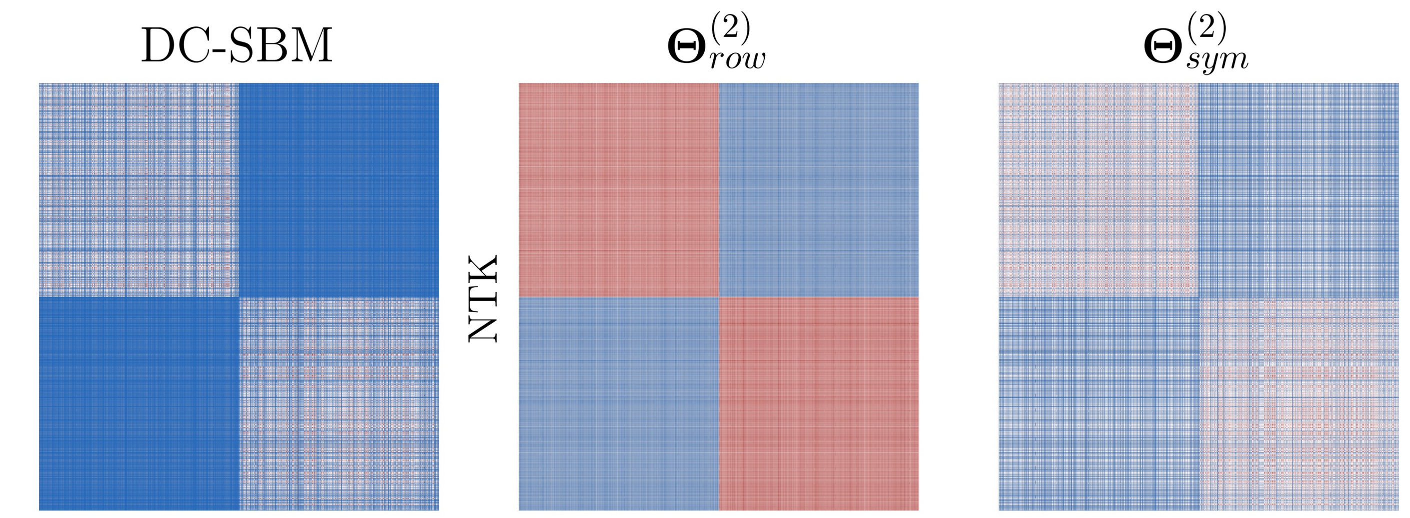
- ❖ Assume $A = M$
- ❖ Compute GNTK with $A = M$
- ❖ Measure class separability of the kernel

**Class sep. $\zeta(\Theta^{(d)})$ = avg. in-class and out-of-class block difference**

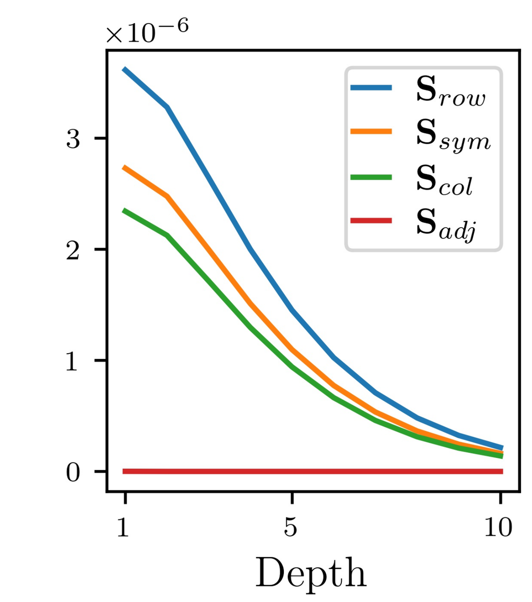
**Larger $\zeta(\Theta^{(d)})$ → better preservation of the block structure**

## Results

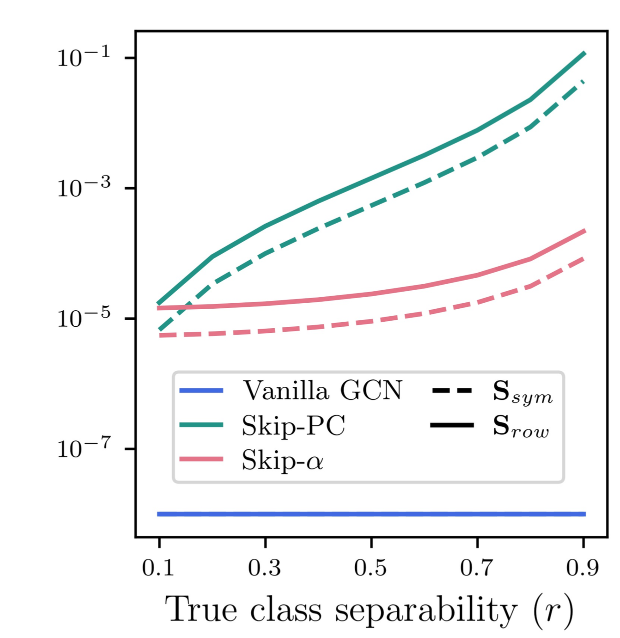
$\zeta(\Theta^{(d)})$ of $S_{row}$ > $S_{sym}$



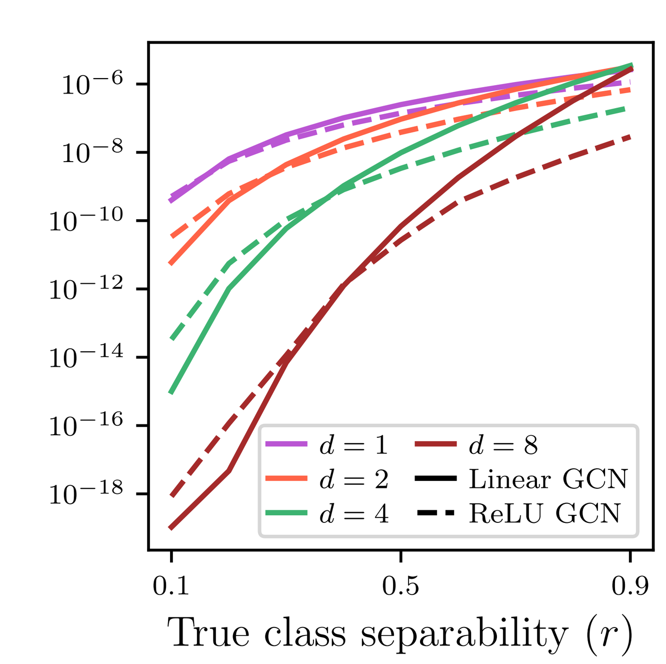

$\zeta(\Theta^{(d)})$ ↓ as $d$ ↑



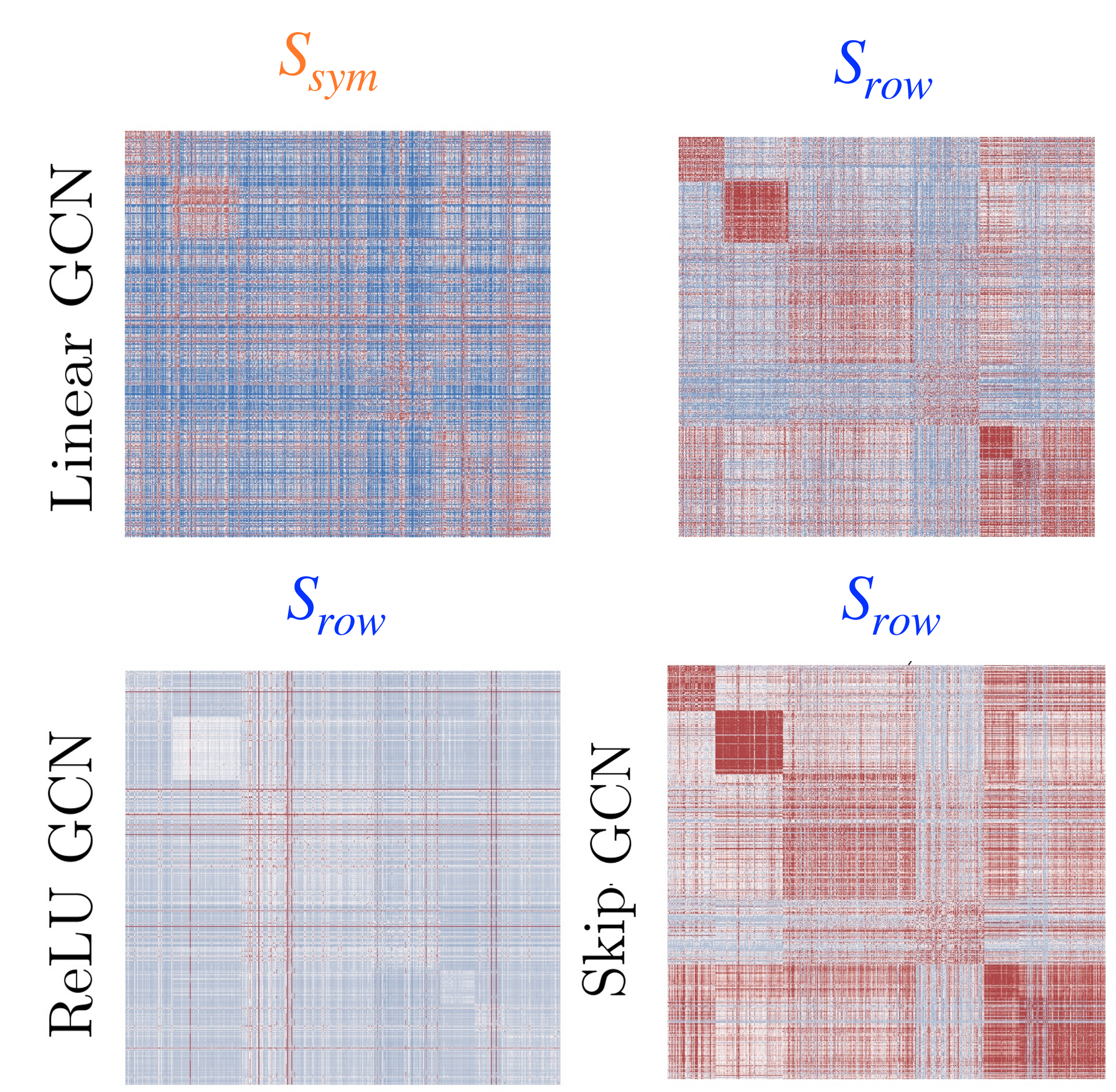

Skip connection
$\zeta(\Theta^{(\infty)})$ is $\mathcal{O}(r^2)$




Linear vs ReLU
$\zeta$(linear) ≃ $\zeta$(relu)




Analysis of real data
Cora




depth=8

Check out the paper here!