Foundations of Deep Learning

ALF

Alfredo Canziani

@alfcnz

Graph Convolutional Networks

Exploiting domain sparsity

$$\boldsymbol{h} = \boldsymbol{X}\boldsymbol{a}$$

Self-attention (I)

$$\{\boldsymbol{x}_i\}_{i=1}^{t} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \boldsymbol{x}_t\} \rightsquigarrow \boldsymbol{X} \in \mathbb{R}^{n \times t}, \quad \boldsymbol{x}_i \in \mathbb{R}^n$$

$$\boldsymbol{h} = \alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \cdots + \alpha_t \boldsymbol{x}_t = \boldsymbol{X}\boldsymbol{a} \in \mathbb{R}^n$$

$$\alpha_i > 0$$

$$\boldsymbol{X} \doteq \begin{bmatrix} | & | & & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_t \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times t}$$

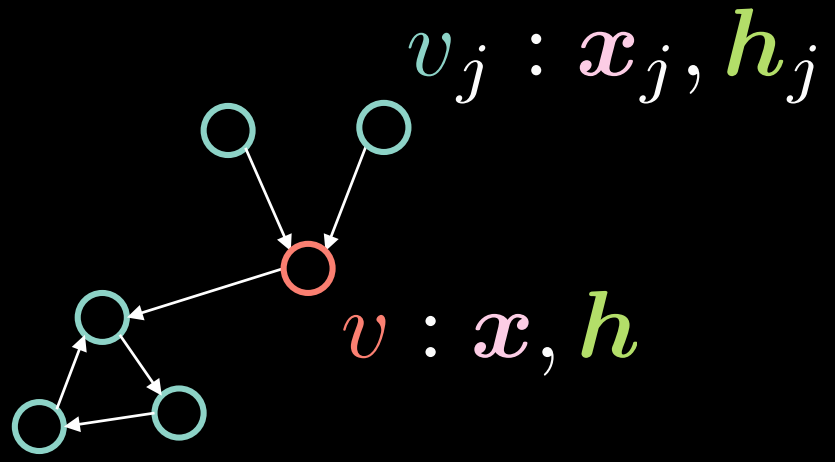
soft attention: $\|\boldsymbol{a}\|_1 = 1$

hard attention: $\|\boldsymbol{a}\|_0 = 1$

GCN

$\boldsymbol{a}$: adjacency vector

$$\alpha_j \triangleq 1 \Leftrightarrow v_j \to v$$

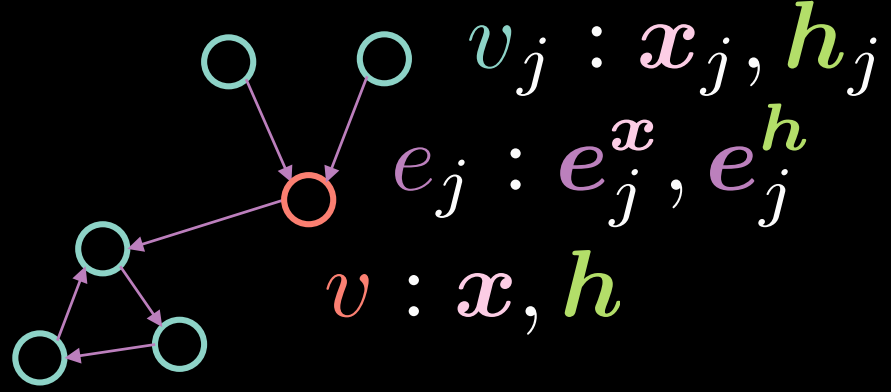


$d = \|\boldsymbol{a}\|_1$: degree (# of incoming edges)

$$\boldsymbol{h} = f(\boldsymbol{U}\boldsymbol{x} + \boldsymbol{V}\boldsymbol{X}\boldsymbol{a}d^{-1}) \qquad f(\cdot) : (\cdot)^+, \sigma(\cdot), \tanh(\cdot)$$

$$\{\boldsymbol{x}_i\}_{i=1}^t \rightsquigarrow \boldsymbol{H} = f(\boldsymbol{U}\boldsymbol{X} + \boldsymbol{V}\boldsymbol{X}\boldsymbol{A}D^{-1}) \quad D = \mathrm{diag}(d_i)$$

Residual gated GCN



$$\boldsymbol{h} = \boldsymbol{x} + \Big(A\boldsymbol{x} + \sum_{v_j \to v} \eta(\boldsymbol{e}_j^{\boldsymbol{h}}) \odot B\boldsymbol{x}_j \Big)^+$$

$$\eta(\boldsymbol{e}_j^{\boldsymbol{h}}) = \sigma(\boldsymbol{e}_j^{\boldsymbol{h}}) \Big(\sum_{v_k \to v} \sigma(\boldsymbol{e}_k^{\boldsymbol{h}}) \Big)^{-1}$$

$$\boldsymbol{e}_j^{\boldsymbol{h}} = \boldsymbol{e}_j^{\boldsymbol{x}} + \Big(C\boldsymbol{e}_j^{\boldsymbol{x}} + D\boldsymbol{x}_j + E\boldsymbol{x} \Big)^+$$

Bresson & Laurent (2018) Residual Gated Graph ConvNets