# Graph Convolutional Networks

Exploiting domain sparsity

from the *attention* lecture

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

# 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: $\|\boldsymbol{a}\|_0 = 1$

from the *natural signals* lecture

# Input layer / samples

$$\mathcal{X} = \{\boldsymbol{x}^{(i)} \in \mathbb{R}^n \mid \boldsymbol{x}^{(i)} \text{ is a data sample}\}_{i=1}^{m}$$ input samples

$$\boxed{\mathcal{X} = \{\boldsymbol{x}^{(i)} : \overset{\text{domain}}{\Omega} \to \overset{\text{channels}}{\mathbb{R}^c}, \boldsymbol{\omega} \mapsto \boldsymbol{x}^{(i)}(\boldsymbol{\omega})\}_{i=1}^{m}}$$

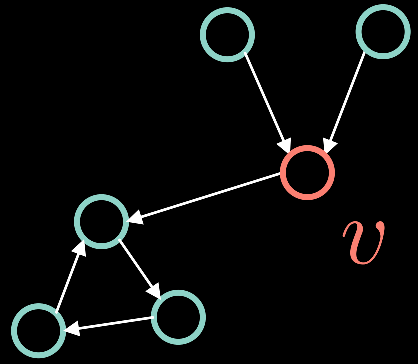
$$\Omega = \{1, 2, \cdots, \overset{\text{total time}}{T} / \underset{\text{sampling interval}}{\Delta t}\} \subset \mathbb{N}, \quad c \in \{\underset{\text{mono}}{1}, \overset{\text{stereo}}{2}, \underset{\text{Dolby 5.1}}{5+1}, \cdots\}$$

$$\Omega = \{1, \cdots, \overset{\text{height}}{h}\} \times \{1, \cdots, \overset{\text{width}}{w}\} \subset \mathbb{N}^2, \quad c \in \{\overset{\text{grey scale}}{1}, \underset{\text{colour}}{3}, \underset{\text{hyperspectral}}{20}, \cdots\}$$

$$\Omega = \underset{\text{space-time}}{\mathbb{R}^4} \times \overset{\text{four-momentum}}{\mathbb{R}^4}, \quad c = \underset{\text{Hamiltonian}}{1}$$

$$\boldsymbol{x}(\omega_1, \omega_2) = \begin{pmatrix} r(\omega_1, \omega_2) \\ g(\omega_1, \omega_2) \\ b(\omega_1, \omega_2) \end{pmatrix}$$

# GCN




$\boldsymbol{a}$ : adjacency vector

$$\alpha_j \overset{\downarrow}{=} 1 \Leftrightarrow v_j \to v$$

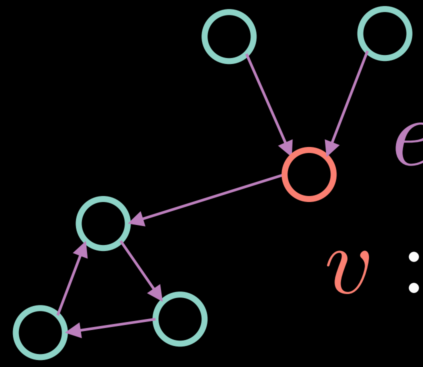
$d = \boldsymbol{a}^\top \mathbf{1}$ : degree (# of incoming edges)

$$\boldsymbol{h} = f(\boldsymbol{U}\boldsymbol{x} + \boldsymbol{V}\boldsymbol{X}\boldsymbol{a}d^{-1})$$

$$f(\cdot) : (\cdot)^+, \sigma(\cdot), \tanh(\cdot)$$

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

# Residual gated GCN



$v_j : \boldsymbol{x}_j, \boldsymbol{h}_j$

$e_j : \boldsymbol{e}_j^{\boldsymbol{x}}, \boldsymbol{e}_j^{\boldsymbol{h}}$

$v : \boldsymbol{x}, \boldsymbol{h}$

$$\boldsymbol{h}^{\ell+1} \to \boldsymbol{h} = \overset{\boldsymbol{h}^{\ell}}{\overset{\downarrow}{\boldsymbol{x}}} + \left( A\boldsymbol{x} + \sum_{v_j \to v} \eta(\boldsymbol{e}_j) \odot B \underset{\substack{\uparrow \\ \boldsymbol{h}_j^{\ell}}}{\boldsymbol{x}_j} \right)^{+}$$

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

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

Bresson & Laurent (2018) Residual Gated Graph ConvNets