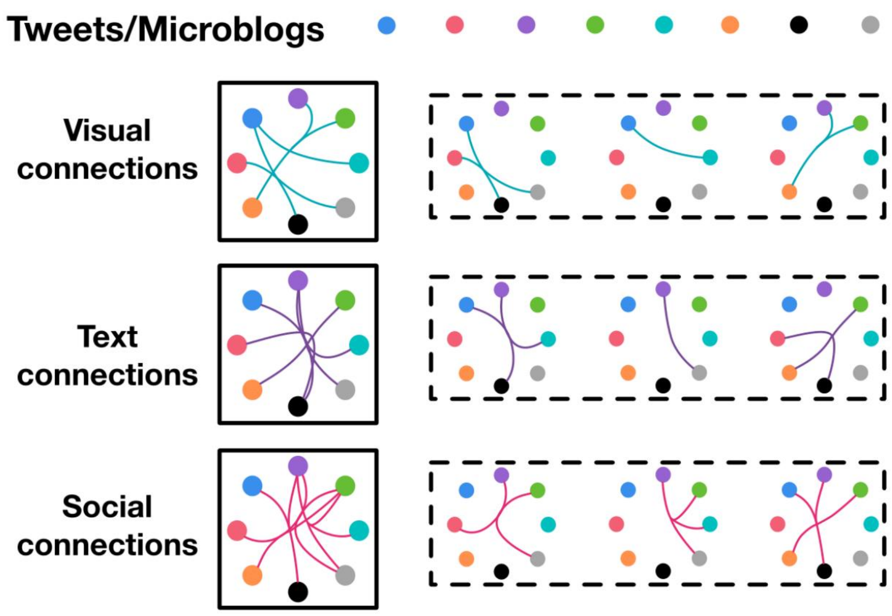
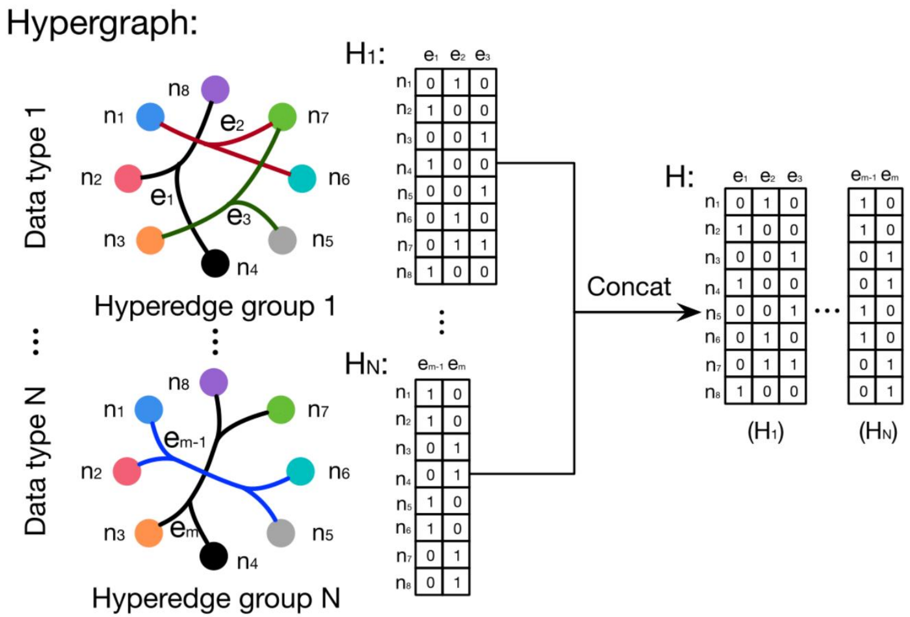
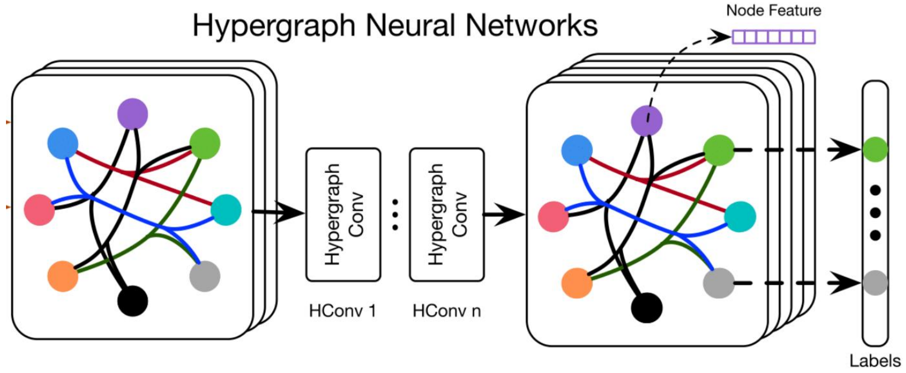
Hypergraph Neural Networks

Presenter: Jiahao Lai

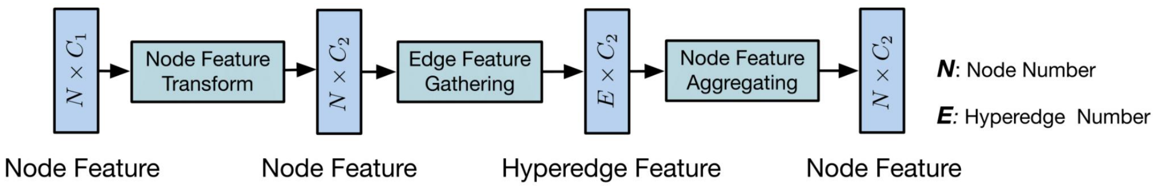
Framework







$$\mathbf{X}^{(l+1)} = \sigma(\mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^\top\mathbf{D}_v^{-1/2}\mathbf{X}^{(l)}\mathbf{\Theta}^{(l)})$$



Feature	Features for Structure					
	GVCNN		MVCNN		GVCNN+MVCNN	
	GCN	HGNN	GCN	HGNN	GCN	HGNN
GVCNN ((Feng et al. 2018))	78.8%	**82.5%**	78.8%	**79.1%**	75.9%	**84.2%**
MVCNN ((Su et al. 2015))	74.0%	**77.2%**	71.3%	**75.6%**	73.2%	**83.6%**
GVCNN+MVCNN	–	–	–	–	76.1%	**84.2%**

Table 5: Comparison between GCN and HGNN on the NTU dataset.

Graph Signal Processing

- Laplacian

$$L = D - A$$

$$\mathcal{L} = U \Lambda U^{-1} = U \Lambda U^T$$

- Graph Signal Variance

$$f^T L f = f^T D f - f^T W f = \sum_{i=1}^{n} D_{i,i} f_i^2 - \sum_{i,j}^{n} f_i f_j W_{i,j}$$

$$= \frac{1}{2}\left(\sum_{i=1}^{n} D_{i,i} f_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} f_i f_j W_{i,j} + \sum_{j=1}^{n} D_{j,j} f_j^2\right)$$

$$= \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} W_{i,j}(f_i - f_j)^2\right) \geq 0$$

- Normalized Laplacian

$$\mathcal{L} = I - D^{-1} A$$

$$\mathcal{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2},$$

- Incidence Matrix ∇ |E| × |V|

$$\nabla_{i,j} = \begin{cases} \nabla_{i,j} = -1 & \text{if } v_j \text{ is the initial vertex of edge } e_i \\ \nabla_{i,j} = 1 & \text{if } v_j \text{ is the terminal vertex of edge of } e_i \\ \nabla_{i,j} = 0 & \text{if } v_j \text{ is not in } e_i \end{cases}$$

$$\nabla \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{bmatrix} = \begin{bmatrix} f_4 - f_1 \\ f_1 - f_2 \\ f_2 - f_4 \\ f_4 - f_3 \\ f_5 - f_4 \\ f_3 - f_5 \end{bmatrix}$$

$$\nabla^T(\nabla f) = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} f_4 - f_1 \\ f_1 - f_2 \\ f_2 - f_4 \\ f_4 - f_3 \\ f_5 - f_4 \\ f_3 - f_5 \end{bmatrix} = \begin{bmatrix} 2f_1 - f_2 - f_4 \\ 2f_2 - f_1 - f_4 \\ 2f_3 - f_4 - f_5 \\ 4f_4 - f_1 - f_2 - f_3 - f_5 \\ 2f_5 - f_3 - f_4 \end{bmatrix}$$

$$\nabla^2 f = [f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1)] - 4f(x, y)$$

$$Lf = \nabla^T(\nabla f)$$

$$(Lf)_i = \sum_{j=1}^{n} W_{i,j}(f_i - f_j),$$

Graph Signal Processing

- Graph Fourier Transform

$$\mathcal{F}[\mathrm{f(t)}] = \int \mathrm{f(t)e^{-i\omega t}dt}$$

$$\mathscr{F}[(\mathrm{f} * \mathrm{g})(\mathrm{t})] = \mathscr{F}(\mathrm{f(t)}) \odot \mathscr{F}(\mathrm{g(t)})$$

$$(\mathrm{f} * \mathrm{g})(\mathrm{t}) = \mathscr{F}^{-1}[\mathscr{F}(\mathrm{f(t)}) \odot \mathscr{F}(\mathrm{g(t)})]$$

$$\mathcal{L} = U\Lambda U^{-1} = U\Lambda U^{T}$$

$$\mathscr{F}(\mathbf{f}) = \mathbf{U}^{\mathrm{T}}\mathbf{f}$$

$$\mathscr{F}^{-1}(\hat{\mathbf{f}}) = \mathbf{U}\hat{\mathbf{f}}$$

- Graph Filter

$$\begin{aligned} \mathbf{f} *_{\mathrm{G}} \mathbf{g} &= \mathscr{F}^{-1}(\mathscr{F}(\mathbf{f}) \odot \mathscr{F}(\mathbf{g})) \\ &= \mathbf{U}\left(\mathbf{U}^{\mathrm{T}}\mathbf{f} \odot \mathbf{U}^{\mathrm{T}}\mathbf{g}\right) \\ &= \mathbf{U}\left(\hat{\mathbf{f}} \odot \hat{\mathbf{g}}\right) \\ &= \mathbf{U}\mathbf{g}_{\theta}(\mathbf{\Lambda})\mathbf{U}^{\mathrm{T}}\mathbf{f} \end{aligned}$$

- Spectral CNN

$$g_{\theta}{}_{i,j}^{(k)} = \begin{bmatrix} \theta_1^{(k)} & & & \\ & \theta_2^{(k)} & & \\ & & \ddots & \\ & & & \theta_n^{(k)} \end{bmatrix}$$

$$f_j^{(k+1)} = \sigma(U \sum_{i=1}^{C_k} g_{\theta}{}_{i,j}^{(k)} U^T f_i^{(k)}) = \sigma(U \sum_{i=1}^{C_k} g_{\theta}{}_{i,j}^{(k)} \hat{f}_i^{(k)})$$

- Spatial K-Localization

$$\mathrm{g}_{\theta}(\mathbf{\Lambda}) = \begin{pmatrix} \sum_{\mathrm{i}=0}^{\mathrm{K}} \theta_{\mathrm{i}} \lambda_1^{\mathrm{i}} & & \\ & \ddots & \\ & & \sum_{\mathrm{i}=0}^{\mathrm{K}} \theta_{\mathrm{i}} \lambda_{\mathrm{n}}^{\mathrm{i}} \end{pmatrix} = \sum_{\mathrm{i}=0}^{\mathrm{K}} \theta_{\mathrm{i}} \Lambda^{\mathrm{i}}$$

$$\mathbf{f} *_{\mathrm{G}} \mathbf{g} = \mathrm{U}(\sum_{\mathrm{i}=0}^{\mathrm{K}} \theta_{\mathrm{i}} \Lambda^{\mathrm{i}})\mathrm{U}^{\mathrm{T}}\mathbf{f} = \sum_{\mathrm{i}=0}^{\mathrm{K}} \theta_{\mathrm{i}} (\mathrm{U}\Lambda^{\mathrm{i}}\mathrm{U}^{\mathrm{T}})\mathbf{f} = \sum_{\mathrm{i}=0}^{\mathrm{K}} \theta_{\mathrm{i}} \mathrm{L}^{\mathrm{i}}\mathbf{f}$$

Graph Signal Processing

- ChebNet

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$$

$$T_0(x) = 1, T_1 = x$$

$$\widetilde{\Lambda} = \frac{2\Lambda}{\lambda_{max}} - I_n \in [-1, 1]$$

$$g_\theta(\Lambda) = \sum_{k=0}^{K-1} \theta_k T_k(\widetilde{\Lambda})$$

$$g_\theta(\Lambda) = \begin{bmatrix} \hat{g}(\lambda_1) & & & \\ & \hat{g}(\lambda_2) & & \\ & & \ddots & \\ & & & \hat{g}(\lambda_n) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=0}^{K-1} \theta_k T_k(\hat{\lambda_1}) & & & \\ & \sum_{k=0}^{K-1} \theta_k T_k(\hat{\lambda_2}) & & \\ & & \ddots & \\ & & & \sum_{k=0}^{K-1} \theta_k T_k(\hat{\lambda_n}) \end{bmatrix}$$

$$g_\theta(\mathcal{L})f = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathcal{L}})f$$

$$\begin{aligned} f(\cdot_G)g_\theta &= g_\theta(U\Lambda U^T)f = U\sum_{k=0}^{K-1} \theta_k T_k(\widetilde{\Lambda})U^T f = \sum_{k=0}^{K-1} U\theta_k T_k(\widetilde{\Lambda})U^T f \\ &= \sum_{k=0}^{K-1} U\theta_k (\sum_{c=0}^{k} \alpha_{kc}\widetilde{\Lambda}^k)U^T f = \sum_{k=0}^{K-1} \theta_k (\sum_{c=0}^{k} \alpha_{kc} U\widetilde{\Lambda}^k U^T)f \\ &= \sum_{k=0}^{K} \theta_k (\sum_{c=0}^{k} \alpha_{kc}(U\widetilde{\Lambda}U^T)^k)f = \sum_{k=0}^{K-1} \theta_k T_k(U\widetilde{\Lambda}U^T)f \\ &= \sum_{k=0}^{K-1} \theta_k T_k(\widetilde{\mathcal{L}})f. \end{aligned}$$

- Spectral CNN versus ChebNet

1. Eigenvector computation is time-consuming.
2. Learnable parameters need attention.
3. Laplace matrix maybe not eigen-decomposed.

Graph Signal Processing

- GCN

$$g_\theta(\Lambda) = \sum_{k=0}^{1} \theta_k T_k(\widetilde{\Lambda})$$

$$\widetilde{\mathcal{L}} = \frac{2}{\lambda_{max}}\mathcal{L} - I_n = \mathcal{L} - I_n \qquad \lambda_{max} \approx 2$$

$$f(\cdot_G)g = \sum_{k=0}^{1} \theta_k T_k(\widetilde{\mathcal{L}})f = \theta_0 T_0(\widetilde{\mathcal{L}})f + \theta_1 T_1(\widetilde{\mathcal{L}})f$$

$$= (\theta_0 + \theta_1(\mathcal{L} - I_n))f = (\theta_0 - \theta_1(D^{-\frac{1}{2}}AD^{\frac{1}{2}}))f$$

$$= (\theta(D^{-\frac{1}{2}}AD^{-\frac{1}{2}} + I_n))f \qquad \text{set } \theta_0 = -\theta_1 = \theta$$

$$\widetilde{A} = A + I_n$$

$$\widetilde{D}_{i,i} = \sum_{j=1}^{n} \widetilde{A}_{i,j}$$

$$f(\cdot_G)g = \theta \widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} f$$

$$\hat{A} = \widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}}$$

$$Z = softmax(\hat{A}ReLU(\hat{A}fW^{<0>})W^{<1>})$$

- Hypergraph Laplacian

$$\text{diag}(\mathbf{W}) = [w(e_1), w(e_2), \ldots, w(e_{|\mathcal{E}|})]$$

$$\mathbf{H} \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{E}|} \qquad \mathbf{H}(v,e) = \begin{cases} 1 & \text{if } v \in e \\ 0 & \text{if } v \notin e \end{cases}$$

$$\mathbf{D}_e \in \mathbb{R}^{|\mathcal{E}|*|\mathcal{E}|} \qquad \delta(e) = \sum_{v\in\mathcal{V}} \mathbf{H}(v,e)$$

$$\mathbf{D}_v \in \mathbb{R}^{|\mathcal{V}|*|\mathcal{V}|} \qquad d(v) = \sum_{e\in\mathcal{E}} w(e) * \mathbf{H}(v,e)$$

$$\Delta = \mathbf{D}_v - \mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^T$$

$$\Delta = \mathbf{I} - \mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^T\mathbf{D}_v^{-1/2}$$

- Hypergraph Signal Variance

$$\Omega(\mathbf{F}) = \frac{1}{2}\sum_{e\in E}\sum_{\{u,v\}\subseteq e} \frac{w(e)}{\delta(e)} \left(\frac{\mathbf{F}(u)}{\sqrt{d(u)}} - \frac{\mathbf{F}(v)}{\sqrt{d(v)}}\right)^2 = \mathbf{F}^T\Delta\mathbf{F}.$$

$$\underset{\mathbf{F}}{\text{argmin}}\Psi(\mathbf{F}) := \left\{\mathbf{F}^T\Delta\mathbf{F} + \lambda\|\mathbf{F} - \mathbf{Y}\|^2\right\}$$

Graph Signal Processing

- HyperGNN

$$\mathbf{\Delta} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^\top$$

$$\mathbf{g} \star \mathbf{x} = \mathbf{\Phi}((\mathbf{\Phi}^\top \mathbf{g}) \odot (\mathbf{\Phi}^\top \mathbf{x})) = \mathbf{\Phi} g(\mathbf{\Lambda}) \mathbf{\Phi}^\top \mathbf{x}$$

$$g(\mathbf{\Lambda}) = \operatorname{diag}(\mathbf{g}(\lambda_\mathbf{1}), \ldots, \mathbf{g}(\lambda_\mathbf{n}))$$

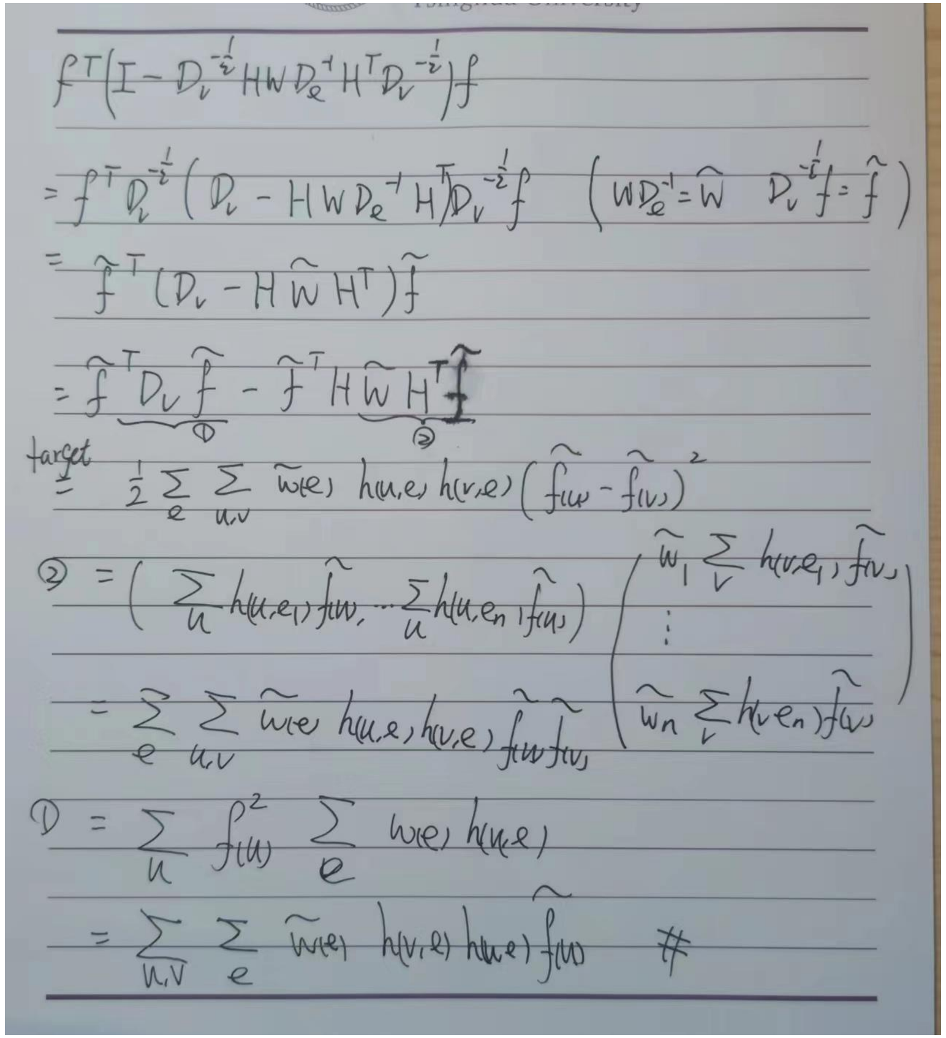
$$\mathbf{g} \star \mathbf{x} \approx \sum_{k=0}^{K} \theta_k T_k(\tilde{\mathbf{\Delta}})\mathbf{x}$$

$$\mathbf{g} \star \mathbf{x} \approx \theta_0 \mathbf{x} - \theta_\mathbf{1} \mathbf{D}^{-\mathbf{1/2}} \mathbf{H} \mathbf{W} \mathbf{D}_\mathbf{e}^{-\mathbf{1}} \mathbf{H}^\top \mathbf{D}_\mathbf{v}^{-\mathbf{1/2}} \mathbf{x}$$

$$\begin{cases} \theta_1 = -\frac{1}{2}\theta \\ \theta_0 = \frac{1}{2}\theta \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \end{cases}$$

$$\begin{aligned} \mathbf{g} \star \mathbf{x} &\approx \frac{1}{2}\theta \mathbf{D}_v^{-1/2} \mathbf{H}(\mathbf{W} + \mathbf{I}) \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \mathbf{x} \\ &\approx \theta \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \mathbf{x}, \end{aligned}$$

$$\mathbf{X}^{(l+1)} = \sigma(\mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \mathbf{X}^{(l)} \mathbf{\Theta}^{(l)})$$



$$f^\top (I - D_v^{-\frac12} H W D_e^{-1} H^\top D_v^{-\frac12}) f$$

$$= f^\top D_v^{-\frac12} (D_v - H W D_e^{-1} H^\top) D_v^{-\frac12} f \quad (W D_e^{-1} = \tilde{W} \quad D_v^{-\frac12} f = \tilde{f})$$

$$= \tilde{f}^\top (D_v - H \tilde{W} H^\top) \tilde{f}$$

$$= \underbrace{\tilde{f}^\top D_v \tilde{f}}_{①} - \underbrace{\tilde{f}^\top H \tilde{W} H^\top \tilde{f}}_{②}$$

target

$$= \frac12 \sum_e \sum_{u,v} \tilde{w}(e) h(u,e) h(v,e) (\tilde{f}(u) - \tilde{f}(v))^2$$

$$② = \left(\sum_u h(u,e_1) \tilde{f}(u), \ldots, \sum_u h(u,e_n) \tilde{f}(u)\right) \begin{pmatrix} \tilde{w}_1 \sum_v h(v,e_1) \tilde{f}(v) \\ \vdots \\ \tilde{w}_n \sum_v h(v,e_n) \tilde{f}(v) \end{pmatrix}$$

$$= \sum_e \sum_{u,v} \tilde{w}(e) h(u,e) h(v,e) \tilde{f}(u) \tilde{f}(v)$$

$$① = \sum_u \tilde{f}^2(u) \sum_e w(e) h(u,e)$$

$$= \sum_{u,v} \sum_e \tilde{w}(e) h(v,e) h(u,e) \tilde{f}(u) \quad \#$$