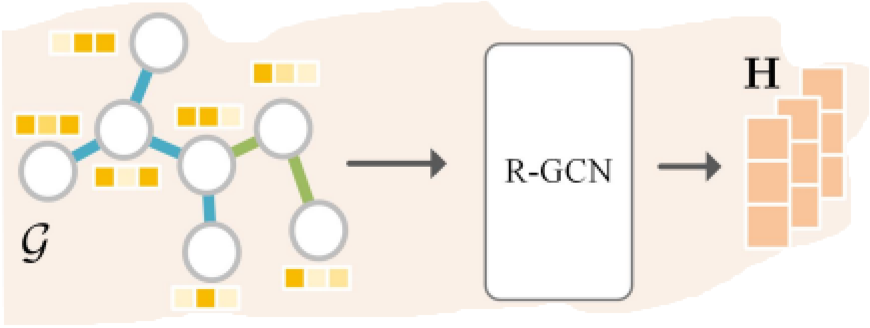
2023_AAAI_Multi-relational Contrastive Learning Graph Neural Network for Drug-drug Interaction Event Prediction



$$\mathbf{h}_v^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{u\in\mathcal{N}_v^r}\frac{1}{c_{vr}}\mathbf{W}_r^{(l)}\mathbf{h}_u^{(l)} + \mathbf{W}_o^{(l)}\mathbf{h}_v^{(l)}\right), \quad (3)$$

$$\mathbf{h}_v = \sum_{l=1}^{L}\alpha_l\mathbf{h}_v^{(l)}, \quad (4)$$

where α_l is a trainable parameter

$\mathbf{h}_v \in \mathbb{R}^Q$. $\mathbf{H} \in \mathbb{R}^{|\mathcal{V}|\times Q}$

R-GCN

where $\sigma(\cdot)$ denotes an activation function, such as ReLU, $\mathcal{N}_v^r$ denotes the set of neighbor nodes of v under relation $r \in \mathcal{R}$, c_{vr} is a problem-specific normalization constant that can either be learned or chosen in advance (such as $c_{vr} = |\mathcal{N}_v^r|$), $\mathbf{h}_v^{(0)} = \mathbf{x}_v$ and $\mathbf{W}_r^{(l)}$ and $\mathbf{W}_o^{(l)}$ denote trainable weight matrices.

2019_EMNLP_Heterogeneous Graph Attention Networks for Semi-supervised Short Text Classification

Heterogeneous Graph Convolution

GCN

$$H^{(l+1)} = \sigma(\tilde{A} \cdot H^{(l)} \cdot W^{(l)}).$$

$$\begin{cases} A' = A + I \\ M_{ii} = \sum_j A'_{ij} \\ \tilde{A} = M^{-\frac{1}{2}} A' M^{-\frac{1}{2}} \end{cases}$$

A: adjacency matrix

HIN

$$H^{(l+1)} = \sigma(\sum_{\tau \in \mathcal{T}} \tilde{A}_\tau \cdot H_\tau^{(l)} \cdot W_\tau^{(l)}),$$

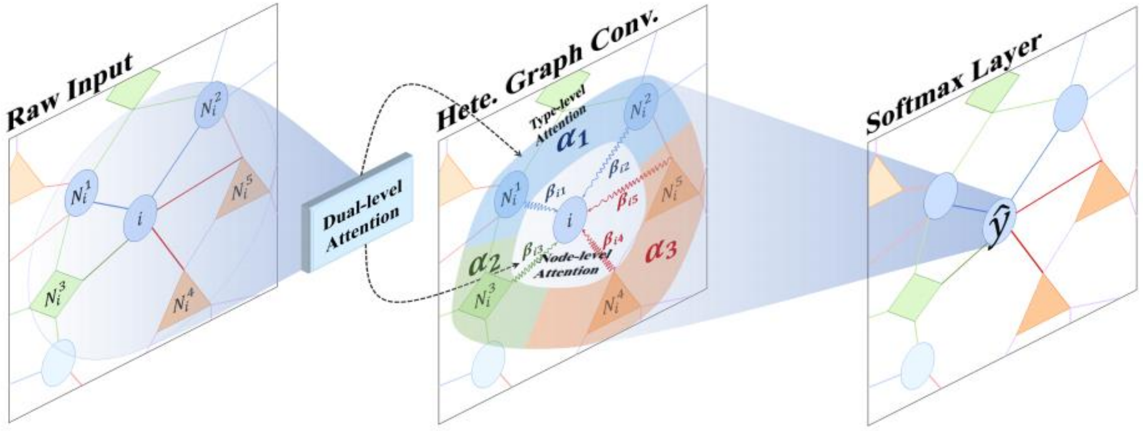
$\mathcal{T}=\{\tau_1, \tau_2, \tau_3\}$ express different types of nodes

$\tilde{A}_\tau \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}_\tau|}$ the submatrix of $\tilde{A}$

$W_\tau^{(l)} \in \mathbb{R}^{q^{(l)} \times q^{(l+1)}}$

Consider the fusion of multiple heterogeneous relationships

2019_EMNLP_Heterogeneous Graph Attention Networks for Semi-supervised Short Text Classification



Dual-level Attention Mechanism

Type-level Attention

$$h_\tau = \sum_{v'} \tilde{A}_{vv'} h_{v'}$$

$$a_\tau = \sigma(\mu_\tau^T \cdot [h_v || h_\tau]),$$

$$\alpha_\tau = \frac{\exp(a_\tau)}{\sum_{\tau' \in \mathcal{T}} \exp(a_{\tau'})}.$$

$\boldsymbol{h_\tau}$: sum of all neighbor nodes feature of node v belong to the same type τ

$\boldsymbol{\alpha_\tau}$:the type-level attention weights

Node-level Attention

$$b_{vv'} = \sigma(\nu^T \cdot \alpha_{\tau'} [h_v || h_{v'}]),$$

$$\beta_{vv'} = \frac{\exp(b_{vv'})}{\sum_{i \in \mathcal{N}_v} \exp(b_{vi})}.$$

$$H^{(l+1)} = \sigma(\sum_{\tau \in \mathcal{T}} \mathcal{B}_\tau \cdot H_\tau^{(l)} \cdot W_\tau^{(l)}).$$

2021_ACL_Directed Acyclic Graph Network for Conversational Emotion Recognition

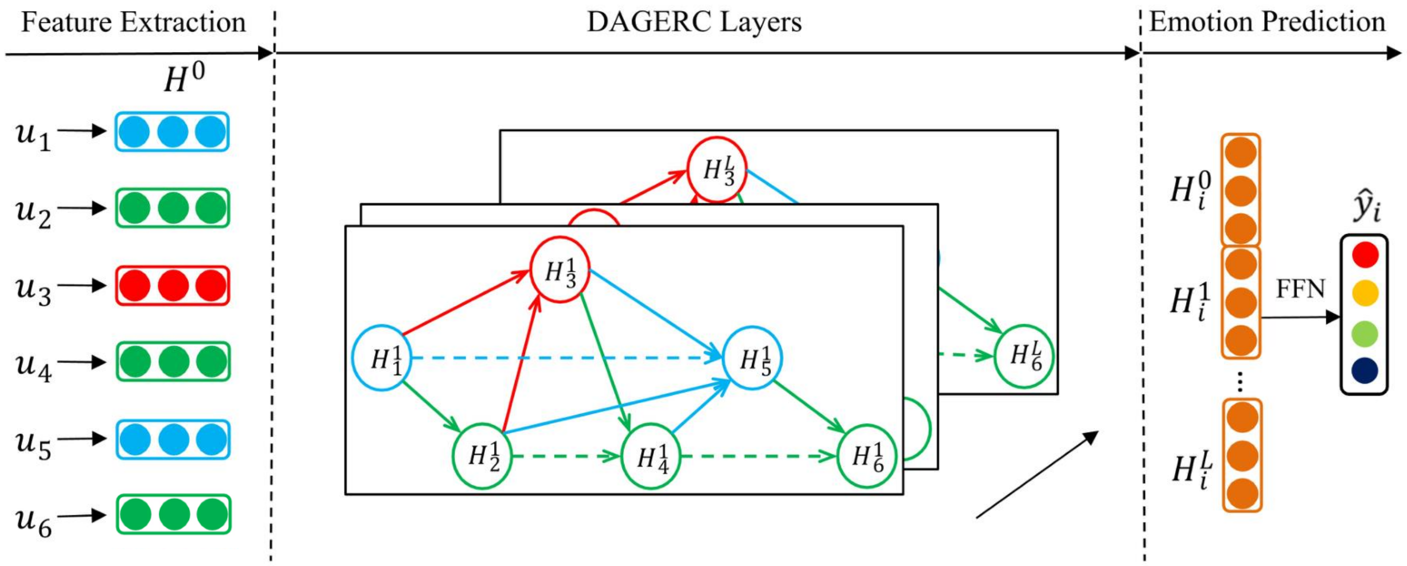


Figure 3: The framework of Directed Acyclic Graph Neural Network for ERC (DAG-ERC).

DAG-ERC Layers

$$\alpha_{ij}^l = \text{Softmax}_{j \in \mathcal{N}_i}(W_\alpha^l [H_j^l \| H_i^{l-1}]) \quad (4)$$

$$M_i^l = \sum_{j \in \mathcal{N}_i} \alpha_{ij} W_{r_{ij}}^l H_j^l,$$

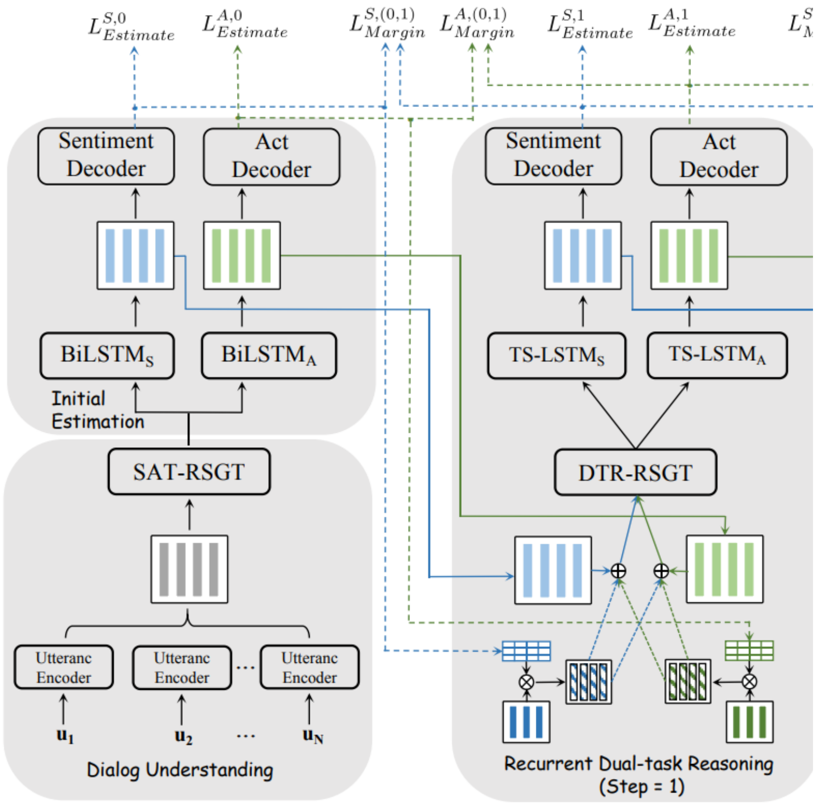
where $W_{r_{ij}}^l \in \{W_0^l, W_1^l\}$ are trainable parameters for the relation-aware transformation.

$$\widetilde{H}_i^l = \text{GRU}_H^l(H_i^{l-1}, M_i^l), \quad (6)$$

$$C_i^l = \text{GRU}_M^l(M_i^l, H_i^{l-1}). \quad (7)$$

$$H_i^l = \widetilde{H}_i^l + C_i^l. \quad (8)$$

2022_ACL_DARER：Dual-task Temporal Relational Recurrent Reasoning Network for Joint Dialog Sentiment Classification and Act Recognition



r_{ij}	1	2	3	4	5	6	7	8
$I_s(i)$	1	1	1	1	2	2	2	2
$I_s(j)$	1	1	2	2	1	1	2	2
$pos(i,j)$	$>$	$\leq$	$>$	$\leq$	$>$	$\leq$	$>$	$\leq$

Table 2: All relation types in SATG (assume there are two speakers). $I_s(i)$ indicates the speaker node i is from. $pos(i,j)$ indicates the relative position of node i and j.

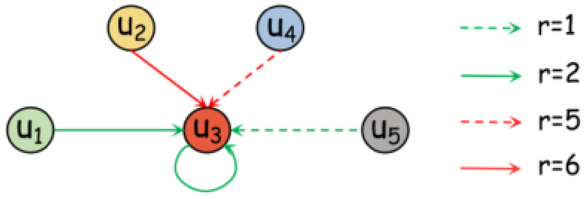


Figure 2: An example of SATG. u_1, u_3 and u_5 are from speaker 1 while u_2 and u_4 are from speaker 2. w.l.o.g, only the edges directed into u_3 node are illustrated.

Speaker-aware Temporal relation-specific graph transformations

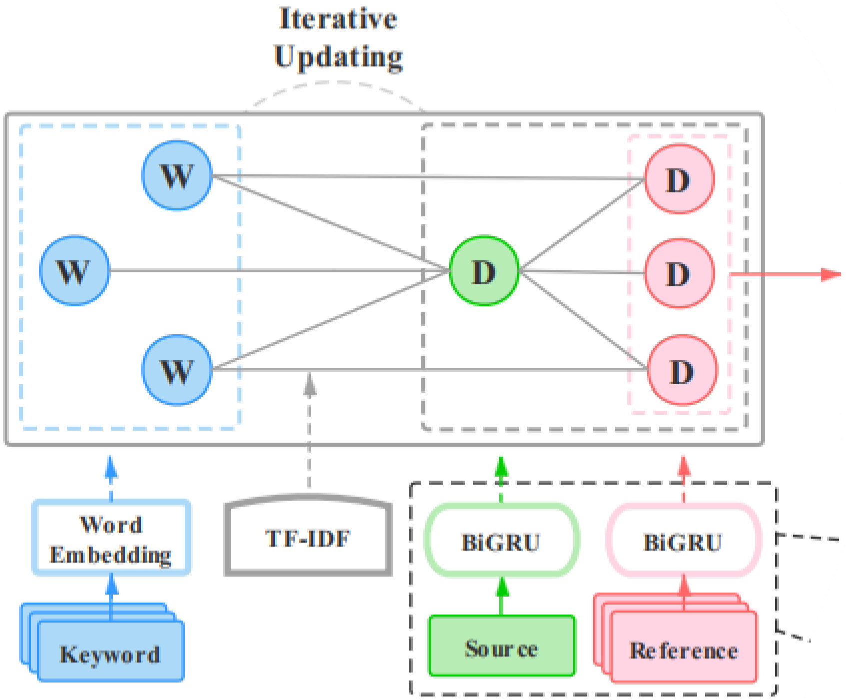
Speaker-aware Temporal RSGT

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$$

$$\boxed{\hat{h}_i = W_1 h_i^0 + \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{|N_i^r|} W_1^r h_j^0 \qquad (1)}$$

$$\hat{H} = (\hat{h}_0, ..., \hat{h}_N)$$

AAAI_2022_Heterogeneous Graph Neural Networks for Keyphrase Generation



Heterogeneous Graph Encoder

$$z_{ij} = \text{LeakyReLU}\left(\mathbf{w}_a^T \left[\mathbf{W}_q \mathbf{h}_i; \mathbf{W}_k \mathbf{h}_j; \mathbf{e}_{ij}\right]\right)$$

$$\alpha_{ij} = \text{softmax}_j\left(z_{ij}\right) = \frac{\exp\left(z_{ij}\right)}{\sum_{k \in \mathcal{N}_i} \exp\left(z_{ik}\right)}$$

$$\mathbf{u}_i = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W}_v \mathbf{h}_j\right),$$

where $\mathbf{e}_{ij}$ is the embedding of edge feature, α_{ij} is the attention weight between $\mathbf{h_i}$ and $\mathbf{h_j}$, and $\mathbf{u}_i$ is the aggregated feature. For simplicity, we will use GAT $(\mathbf{H}, \mathbf{H}, \mathbf{H}, \mathbf{E})$ to denote the GAT aggregating layer, where $\mathbf{H}$ is used for query, key, and value, and $\mathbf{E}$ is used as edge features.

Dual-Dynamic GCN Module

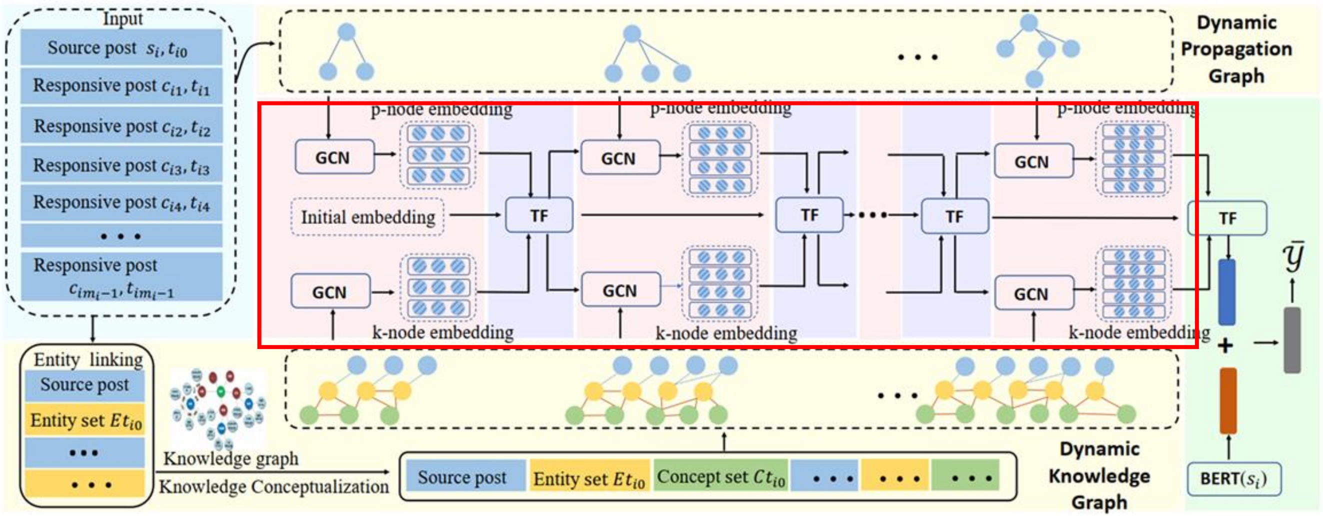
A single Dual-Static GCN unit

$$H_{ir}^{p\,(1)} = g\left(\hat{A}_{ir}^{p} H_{ir}^{p\,(0)} W_r^{p(0)} + b_r^{p(0)}\right) \quad (1)$$

$$H_{ir}^{p\,(2)} = g\left(\hat{A}_{ir}^{p} H_{ir}^{p\,(1)} W_r^{p(1)} + b_r^{p(1)}\right) \quad (2)$$

$$H_{ir}^{p} \in \mathbb{R}^{m_i \times F} \qquad H_{ir}^{p\,(0)} = H_{ir}^{p}$$

$$\hat{A}_{ir}^{p} = (\tilde{D}_{ir}^{p})^{-\frac{1}{2}} \tilde{A}_{ir}^{p} (\tilde{D}_{ir}^{p})^{-\frac{1}{2}} \quad (3)$$



$$\tilde{H}_{ir}^{p\,(l)} = \mathbf{Concat}\left(H_{ir}^{p\,(l)}, (H_{ir}^{p\,(l-1)})^{Source}\right) \quad (4)$$

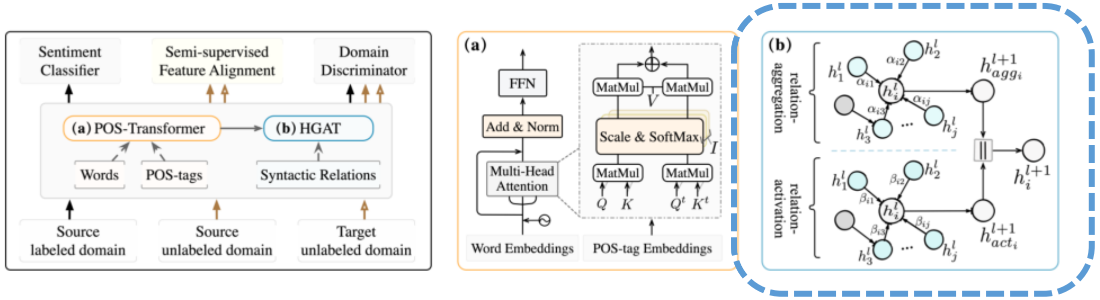
$\tilde{H}_{ir}^{k\,(1)}$ and $\tilde{H}_{ir}^{k\,(2)}$ are obtained in the same manner

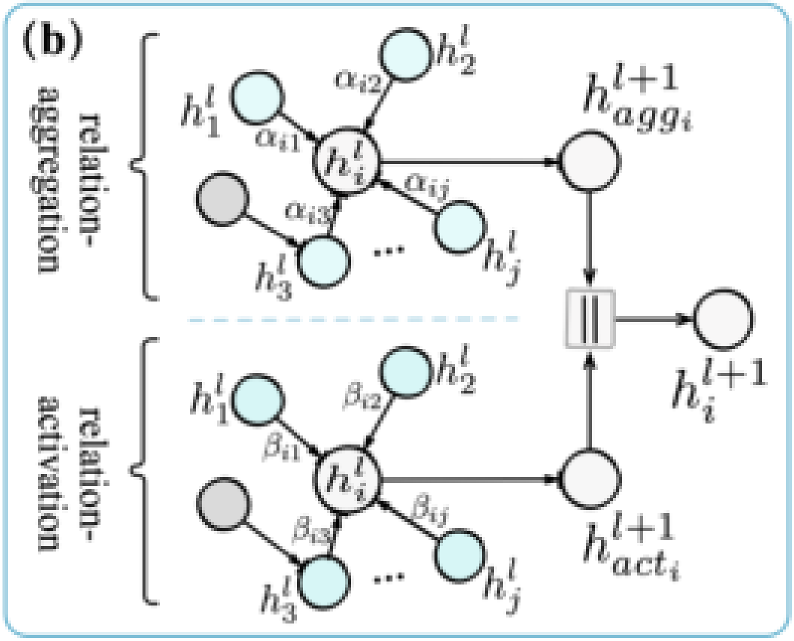
$$H_{ir}^{p\,(2)} = g\left(\hat{A}_{ir}^{p} \tilde{H}_{ir}^{p\,(1)} W_r^{p(1)} + b_r^{p(1)}\right) \quad (5)$$

$$\tilde{H}_{ir}^{p\,(2)} = \mathbf{Concat}\left(H_{i}^{p(2)}, (H_{i}^{p(1)})^{Source}\right) \quad (6)$$

$$\tilde{H}_{ir}^{p\,(2)} = \mathbf{ReLU}(W_r^{p} \tilde{H}_{ir}^{p\,(2)} + b_r^{p}) \quad (7)$$

2022_SIGIR_Graph Adaptive Semantic Transfer for Cross-domain Sentiment Classification





$$h_{agg_i}^{l+1} = \|_{k=1}^{\bar{K}} \sigma\Big(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^{lk} W_{lk} h_j^l \Big), \tag{5}$$

$$f_{ij}^{lk} = \sigma\big(a_{lk}^T [W_{lk} h_i^l \| W_{lk} h_j^l \| \boxed{W_{lk} r_{ij}]}\big), \tag{6}$$

$$\alpha_{ij}^{lk} = \frac{\exp\left(f_{ij}^{lk}\right)}{\sum_{j=1}^{\mathcal{N}_i} \exp\left(f_{ij}^{lk}\right)}, \tag{7}$$

$$\beta_{ij}^{lk} = \frac{\exp\left(F_{act.}(h_i^l, h_j^l)\right)}{\sum_{j=1}^{\mathcal{N}_i} \exp\left(F_{act.}(h_i^l, h_j^l)\right)}, \tag{8}$$

$$F_{act.} = \frac{\left(W_Q^{lk} h_i^l\right)\left(W_K^{lk} h_j^l + W_{Kr}^l r_{ij}\right)^T}{\sqrt{d/\bar{K}}}, \tag{9}$$

参考知识图谱

$$h_{act_i}^{l+1} = \|_{k=1}^{\bar{K}} \sigma\Big(\sum_{j \in \mathcal{N}_i} \beta_{ij}^{lk} \boxed{(W_V^{lk} h_j^l + W_{Vr}^l r_{ij})}\Big), \tag{10}$$

$$h_i^{l+1} = h_{agg_i}^{l+1} \, \| \, h_{act_i}^{l+1}. \tag{11}$$