Learning Conjoint Attentions for Graph Neural Nets

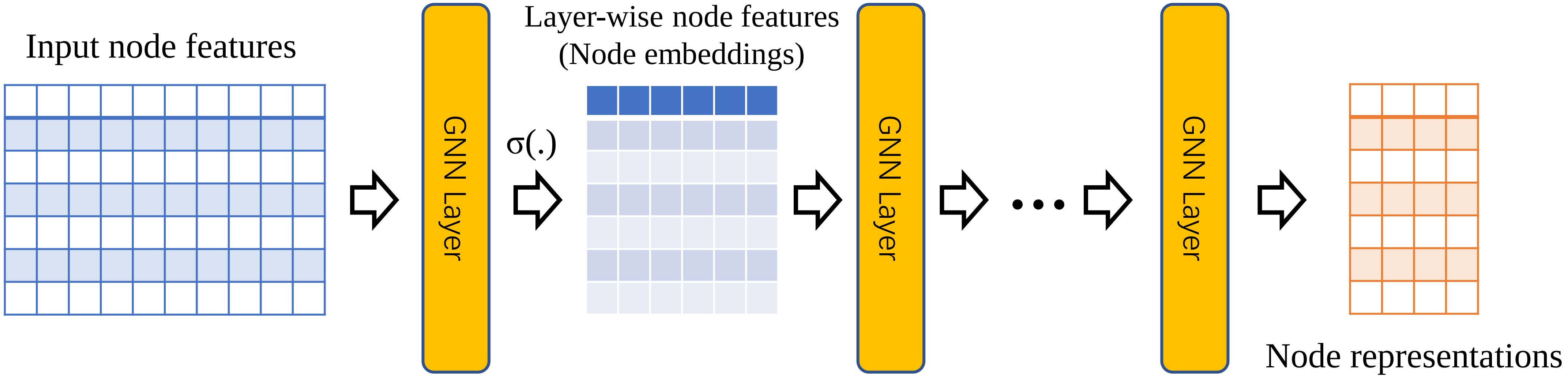
Tiantian He, Yew-Soon Ong, Lu Bai

Agency for Science, Technology and Research (A*STAR)

Nanyang Technological University

NeurIPS 2021

Graph Neural Networks (GNNs)



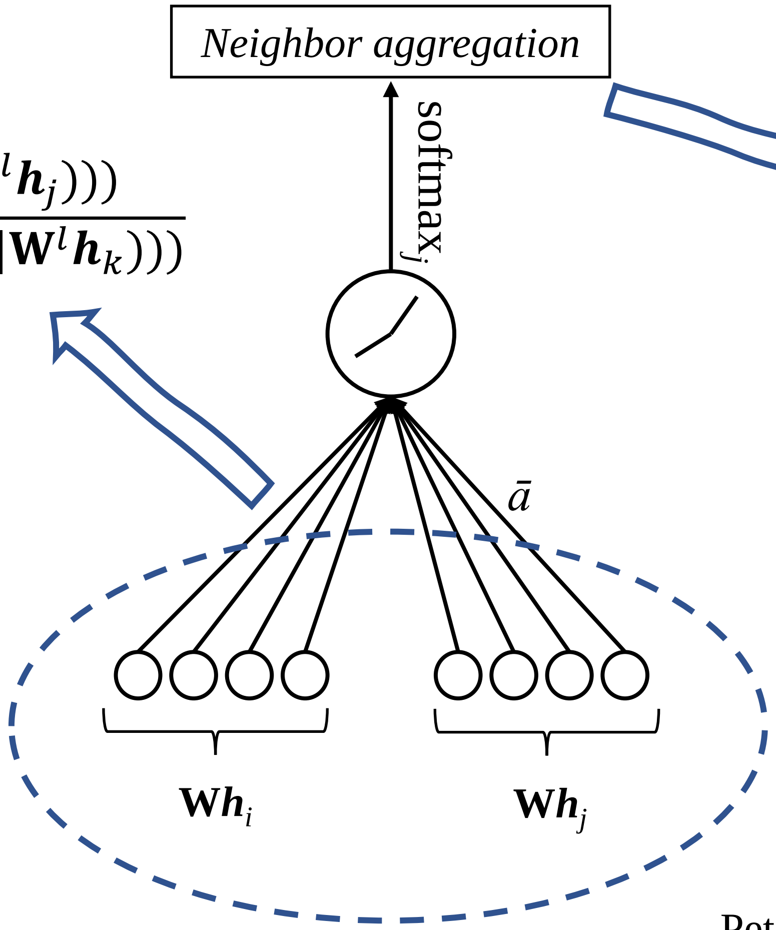
Operation on self features and neighbors'

$$\boldsymbol{h}_i^{l+1} = g((\mathbf{W}^l \boldsymbol{h}_i^l, \mathbf{W}^l \boldsymbol{h}_j^l))_{j \in N_i}$$

- Graph Convolutional Networks
- Graph Attention Networks
- And many others

Classical Graph Attentions

$$\alpha_{ij}^{l} = \frac{\exp(LeakyReLU(\overline{\boldsymbol{a}}(\mathbf{W}^{l}\boldsymbol{h}_i||\mathbf{W}^{l}\boldsymbol{h}_j)))}{\sum_{k\epsilon N_i}\exp(LeakyReLU(\overline{\boldsymbol{a}}(\mathbf{W}^{l}\boldsymbol{h}_i||\mathbf{W}^{l}\boldsymbol{h}_k)))}$$



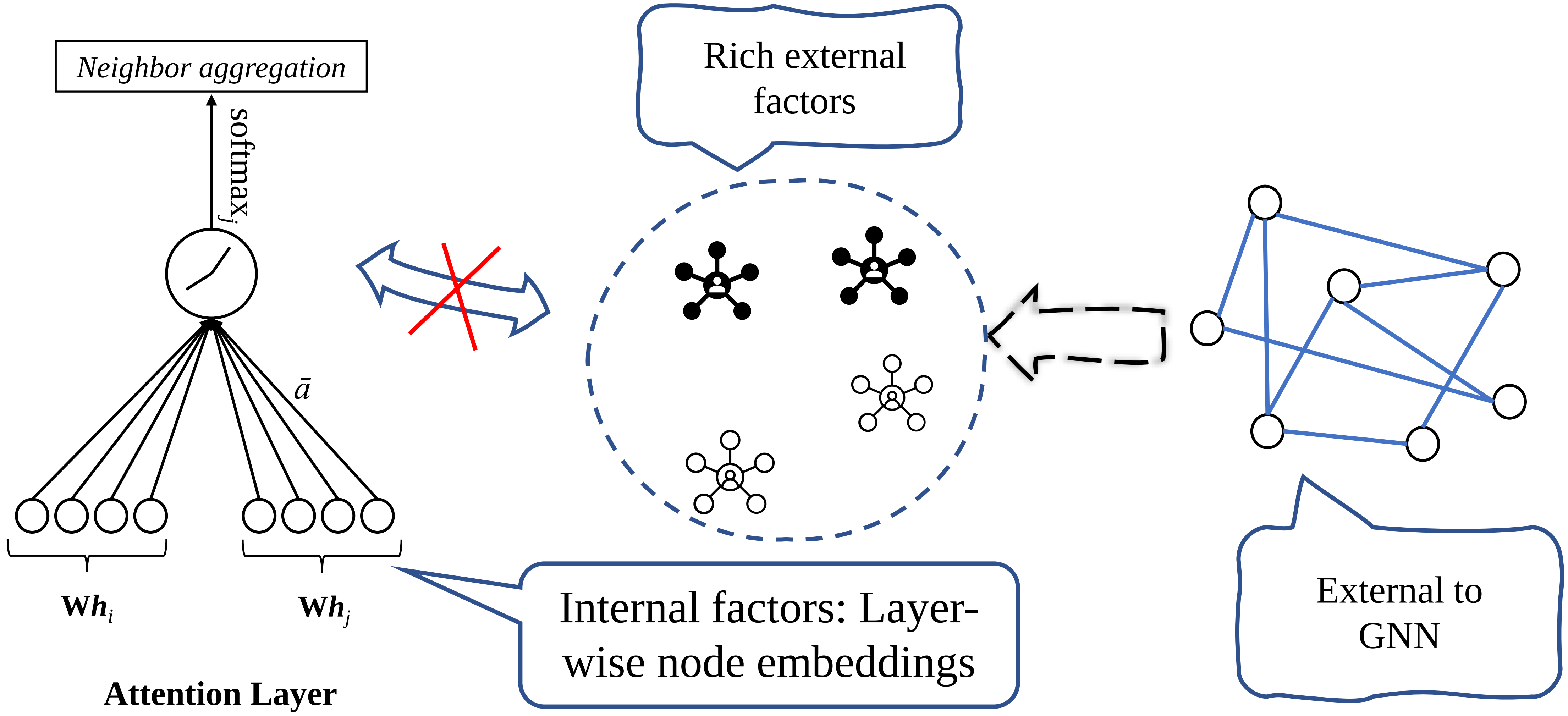
$$\boldsymbol{h}_i^{l+1} = \sum_{j\epsilon N_i} \alpha_{ij}^{l}\mathbf{W}^{l}\boldsymbol{h}_j$$

Internal factors: Layer-wise node embeddings

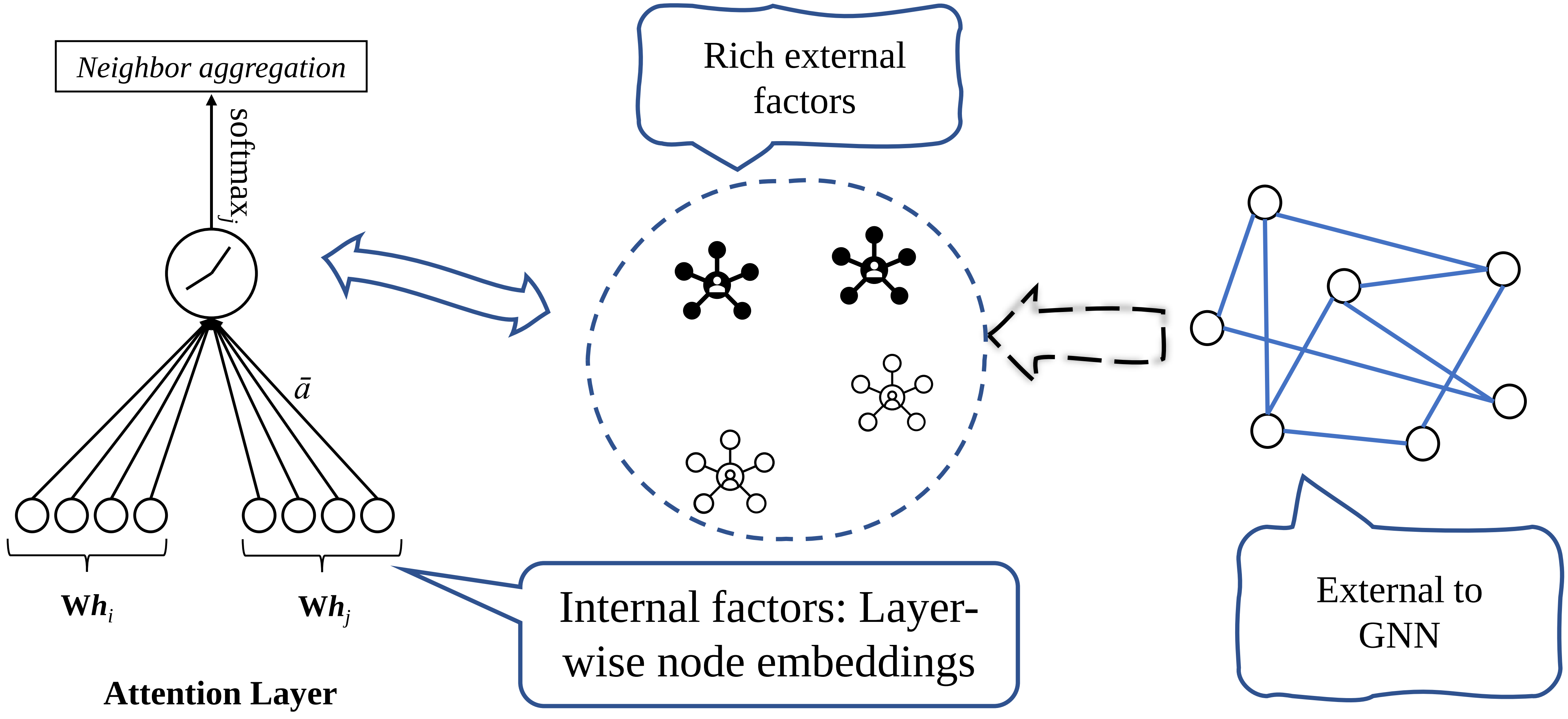
Internal-factor-bias 1-hop neighbor aggregation

Petar Velickovic et. al., 2018, Graph Attention Networks

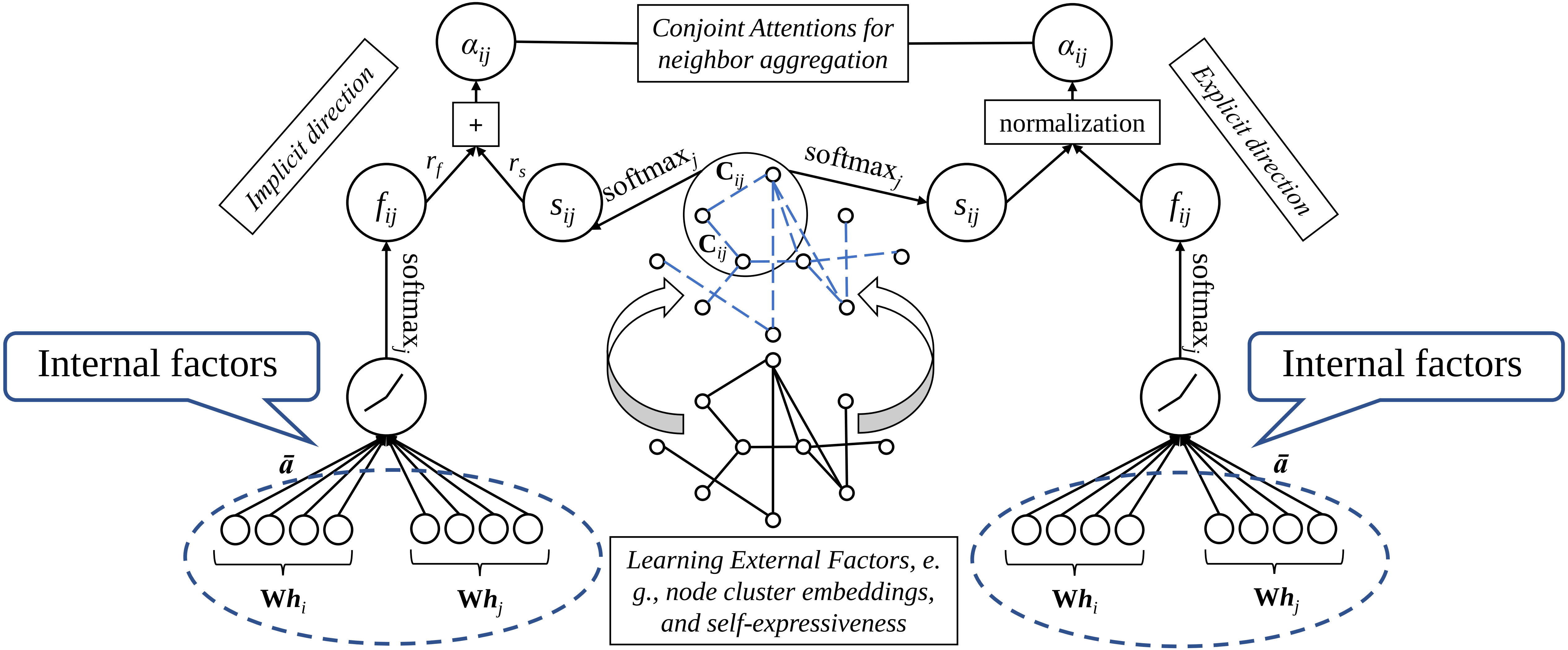
Classical Graph Attentions



Our Approach

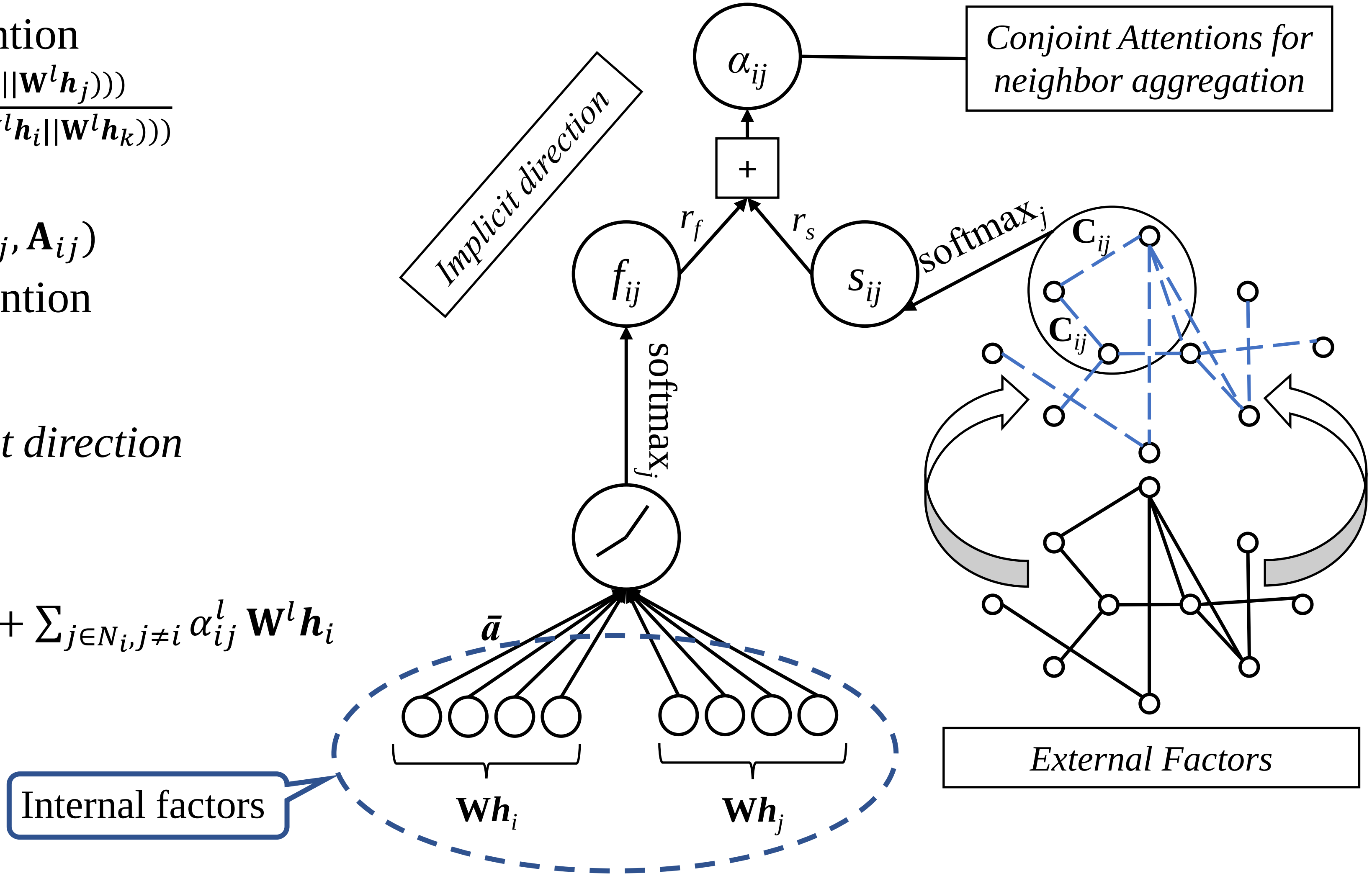


Conjoint Attention Layers (CA Layers)



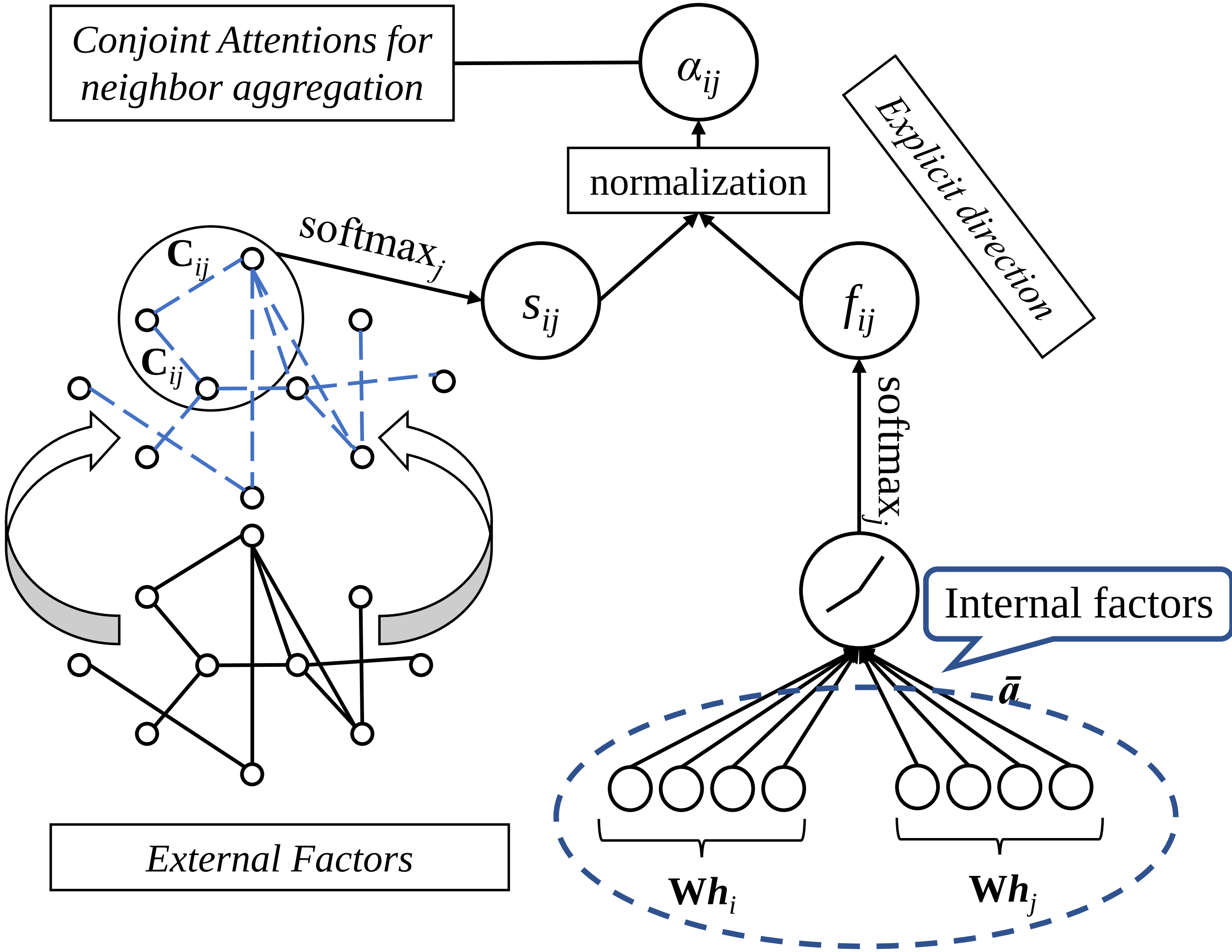
Conjoint Attentions-*Implicit direction*

- Internal factor caused attention
- $f_{ij}^{l} = \frac{\exp(LeakyReLU(\bar{\boldsymbol{a}}(\mathbf{W}^{l}\boldsymbol{h}_{i}||\mathbf{W}^{l}\boldsymbol{h}_{j})))}{\sum_{k\epsilon N_{i}}\exp(LeakyReLU(\bar{\boldsymbol{a}}(\mathbf{W}^{l}\boldsymbol{h}_{i}||\mathbf{W}^{l}\boldsymbol{h}_{k})))}$
- External factor learning
- $\mathbf{C}_{ij} = \arg\min_{\phi(\mathbf{C})_{ij}} \psi(\phi(\mathbf{C})_{ij}, \mathbf{A}_{ij})$
- External factor caused attention
- $s_{ij}^{l} = \frac{\exp(\mathbf{C}_{ij})}{\sum_{k\epsilon N_{i}}\exp(\mathbf{C}_{ik})}$
- Conjoint Attention-*Implicit direction*
- $\alpha_{ij}^{l} = r_f \cdot f_{ij}^{l} + r_s \cdot s_{ij}^{l}$
- Feature aggregation:
- $\boldsymbol{h}_{i}^{l+1} = \left(\alpha_{ii}^{l} + \epsilon \cdot \frac{1}{|N_i|}\right)\mathbf{W}^{l}\boldsymbol{h}_{i} + \sum_{j\in N_i, j\neq i}\alpha_{ij}^{l}\,\mathbf{W}^{l}\boldsymbol{h}_{i}$

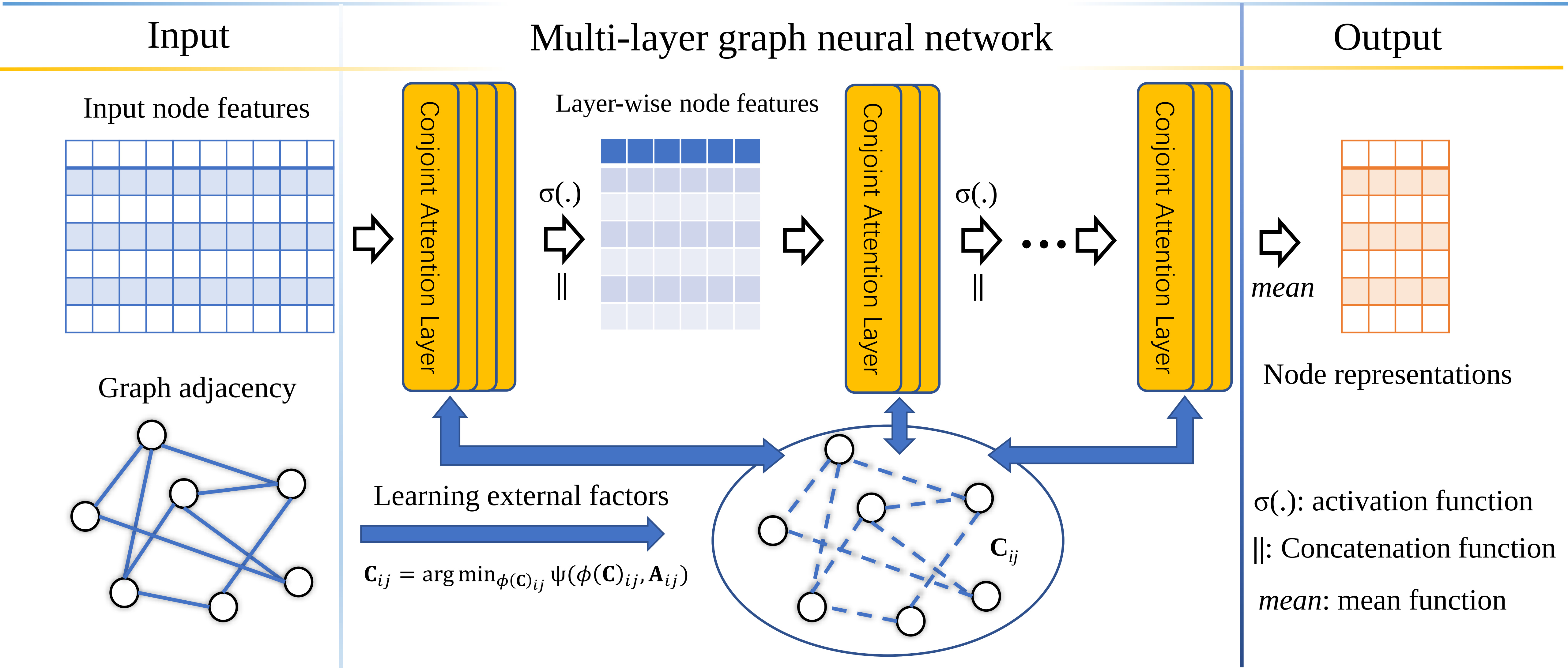


Conjoint Attentions-*Explicit direction*

- Internal factor caused attention
- $f_{ij}^{l} = \frac{\exp(LeakyReLU(\bar{\boldsymbol{a}}(\mathbf{W}^{l}\boldsymbol{h}_i||\mathbf{W}^{l}\boldsymbol{h}_j)))}{\sum_{k\epsilon N_i}\exp(LeakyReLU(\bar{\boldsymbol{a}}(\mathbf{W}^{l}\boldsymbol{h}_i||\mathbf{W}^{l}\boldsymbol{h}_k)))}$
- External factor learning
- $\mathbf{C}_{ij} = \arg\min_{\phi(\mathbf{C})_{ij}} \psi(\phi(\mathbf{C})_{ij}, \mathbf{A}_{ij})$
- External factor caused attention
- $s_{ij}^{l} = \frac{\exp(\mathbf{C}_{ij})}{\sum_{k\epsilon N_i}\exp(\mathbf{C}_{ik})}$
- Conjoint Attention-*Explicit direction*
- $\alpha_{ij}^{l} = \frac{f_{ij}^{l}\cdot s_{ij}^{l}}{\sum_{k\epsilon N_i} f_{ik}^{l}\cdot s_{ik}^{l}}$
- Feature aggregation:
- $\boldsymbol{h}_i^{l+1} = \left(\alpha_{ii}^{l} + \epsilon\cdot\frac{1}{|N_i|}\right)\mathbf{W}^{l}\boldsymbol{h}_i + \sum_{j\in N_i, j\neq i}\alpha_{ij}^{l}\,\mathbf{W}^{l}\boldsymbol{h}_i$



Graph Conjoint Attention Networks (CATs)



Experimental Results

Table 1 Accuracy on Semi-supervised node classification

	Cora	**Citeseer**	**Pubmed**	**CoauthorCS**	**OGB-Arxiv**
GCN	81.42	71.60	79.66	91.54	71.78
GraphSAGE	81.12	71.06	79.04	93.06	69.07
JKNet	78.34	65.88	79.88	89.62	64.91
APPNP	82.80	72.38	82.62	89.16	63.16
GIN	81.58	66.90	80.76	93.03	64.02
GAT	83.84	70.36	81.50	92.80	72.39
GAT-k-Lap	84.10	71.18	82.56	92.70	72.47
CAT-I-MF	85.38	73.22	83.90	93.74	72.89
CAT-I-SC	85.50	73.18	84.28	93.70	72.85
CAT-E-MF	85.56	73.24	83.60	93.40	72.81
CAT-E-SC	85.40	73.02	84.02	93.30	72.83

Experimental Results

Table 2 Accuracy on Semi-supervised node clustering

	Cora	**Citeseer**	**Pubmed**	**CoauthorCS**	**OGB-Arxiv**
GCN	74.25	63.36	77.83	89.74	75.02
GraphSAGE	78.46	69.00	79.52	90.16	73.50
JKNet	75.95	65.12	79.52	86.66	71.28
APPNP	79.93	70.55	82.81	85.93	69.73
GIN	78.25	67.83	79.31	89.97	63.85
GAT	81.39	69.20	80.88	90.09	76.04
GAT-k-Lap	80.66	69.56	81.59	89.83	76.21
CAT-I-MF	82.17	71.15	82.77	90.26	77.72
CAT-I-SC	82.26	71.17	82.86	90.29	77.01
CAT-E-MF	81.98	71.21	82.40	89.66	76.93
CAT-E-SC	82.01	71.11	82.61	89.72	76.98

Thank you very much!

If you are interested in our work, please find more details in our paper.