Meta-Weight Graph Neural Network: Push the Limits Beyond Global Homophily

Xiaojun Ma
mxj@pku.edu.cn
Key Lab. of Machine Perception (MoE), School of AI, Peking University
Beijing, China

Qin Chen
chenqink@pku.edu.cn
Key Lab. of Machine Perception (MoE), School of AI, Peking University
Beijing, China

Yuanyi Ren
celina@pku.edu.cn
Key Lab. of Machine Perception (MoE), School of AI, Peking University
Beijing, China

Guojie Song*
gjsong@pku.edu.cn
Key Lab. of Machine Perception (MoE), School of AI, Peking University
Beijing, China

Liang Wang
liangbo.wl@alibaba-inc.com
Alibaba Inc
Beijing, China

WWW 2022
Code: None

2022.05.07 • ChongQing

Leibniz-Institut für Sozialwissenschaften

Reported by Chenghong Li

Introduction

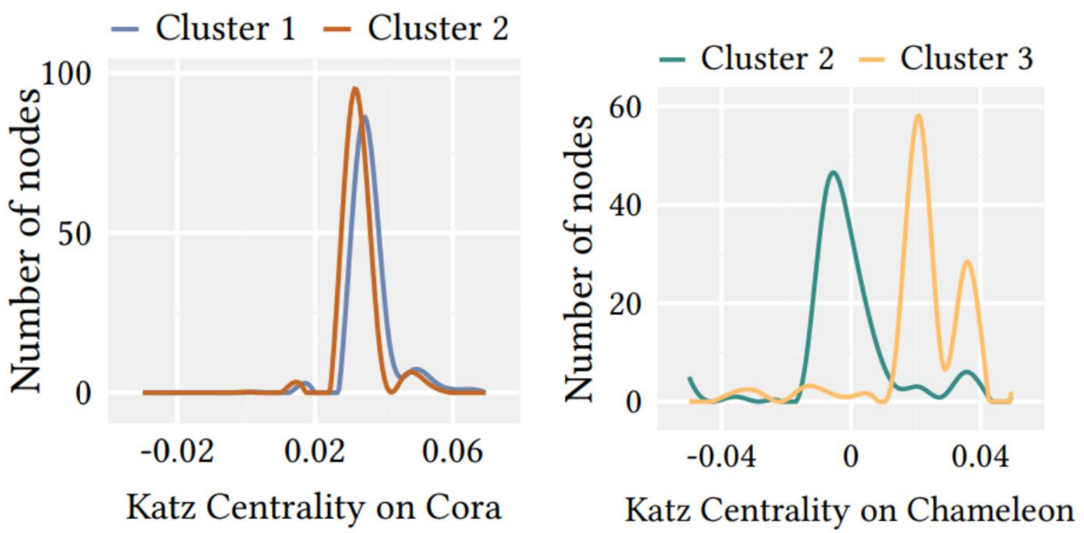


Figure 1: Katz centrality distribution of clustered Cora (in class 1) and Chameleon (in class 4). The nodes are clustered and then the Katz Centrality distribution is plotted for nodes of the same label but belonging in two different clusters.

$$\boldsymbol{H}_i^{(l+1)} = \text{TRANS}\left(\text{AGG}\left(\boldsymbol{H}_i^{(l)}, \left\{\boldsymbol{H}_j^{(l)} : v_j \in N_i\right\}\right)\right), \tag{1}$$

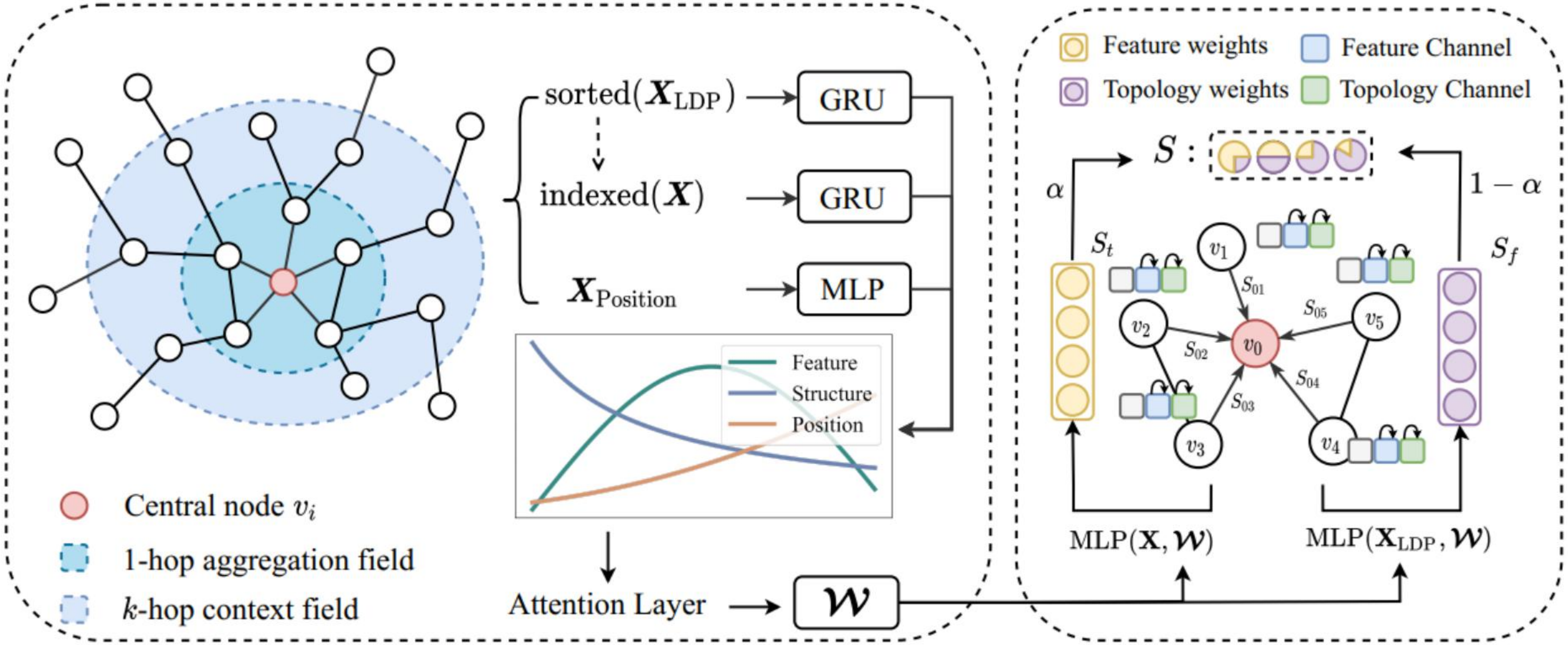
(Global Edge Homophily).

$$h = \frac{|\{(v_i, v_j) : (v_i, v_j) \in \mathcal{E} \wedge \boldsymbol{y}_i = \boldsymbol{y}_j\}|}{|\mathcal{E}|}, \tag{2}$$

(Local Edge Homophily).

$$h_i = \frac{|\{(v_i, v_j) : v_j \in \mathcal{N}_i \wedge \boldsymbol{y}_i = \boldsymbol{y}_j)\}|}{|\mathcal{N}_i|}, \tag{3}$$

Method

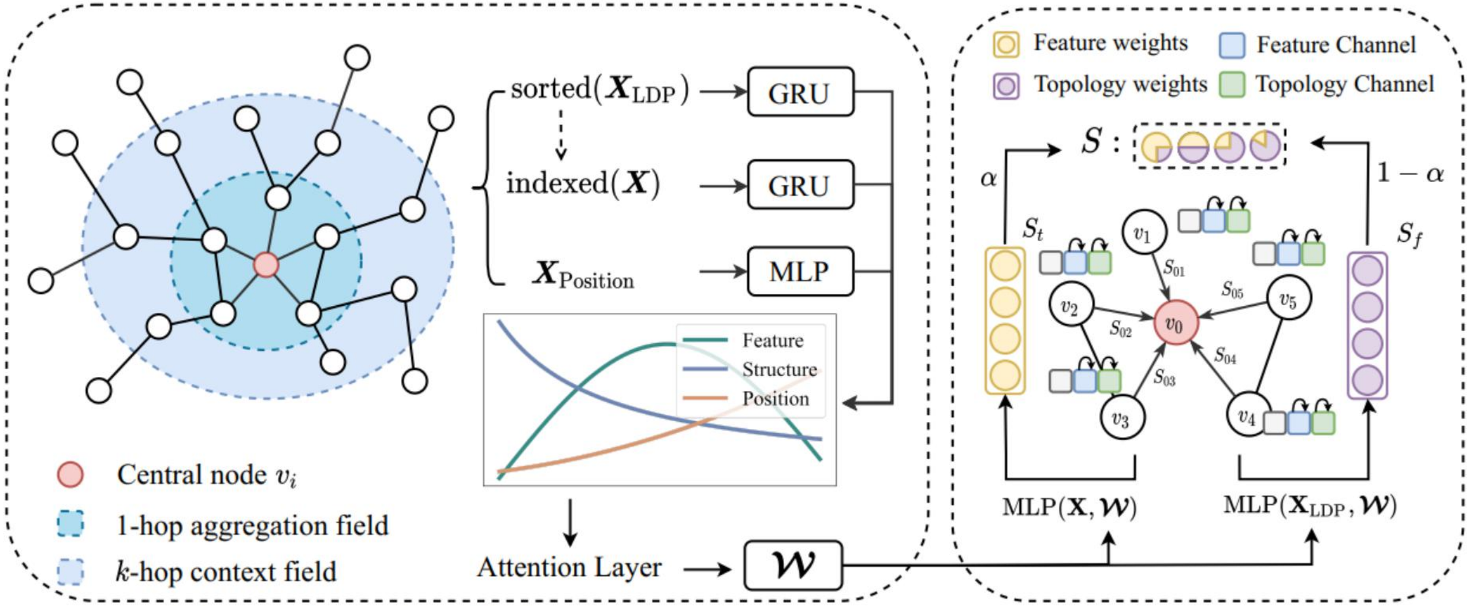


(a) Modeling Node Local Distribution with Meta-Weight

(b) Adaptive Convolution

Figure 2: The framework of MWGNN. (a) Generate the Meta-Weight considering k-hop context field for central nodes. First we learn three local distributions in topological structure, node feature, and positional identity fields and integrate them with an attention layer. (b) Based on the Meta-Weight, we propose the Adaptive Convolution. By generating S_t, S_f and adaptively fusing them with a hyper-parameter α, the Adaptive Convolution aggregates the neighbors. Then two additional Independent Convolution Channels are proposed to boost the node representations efficiently.

Method



(a) Modeling Node Local Distribution with Meta-Weight (b) Adaptive Convolution

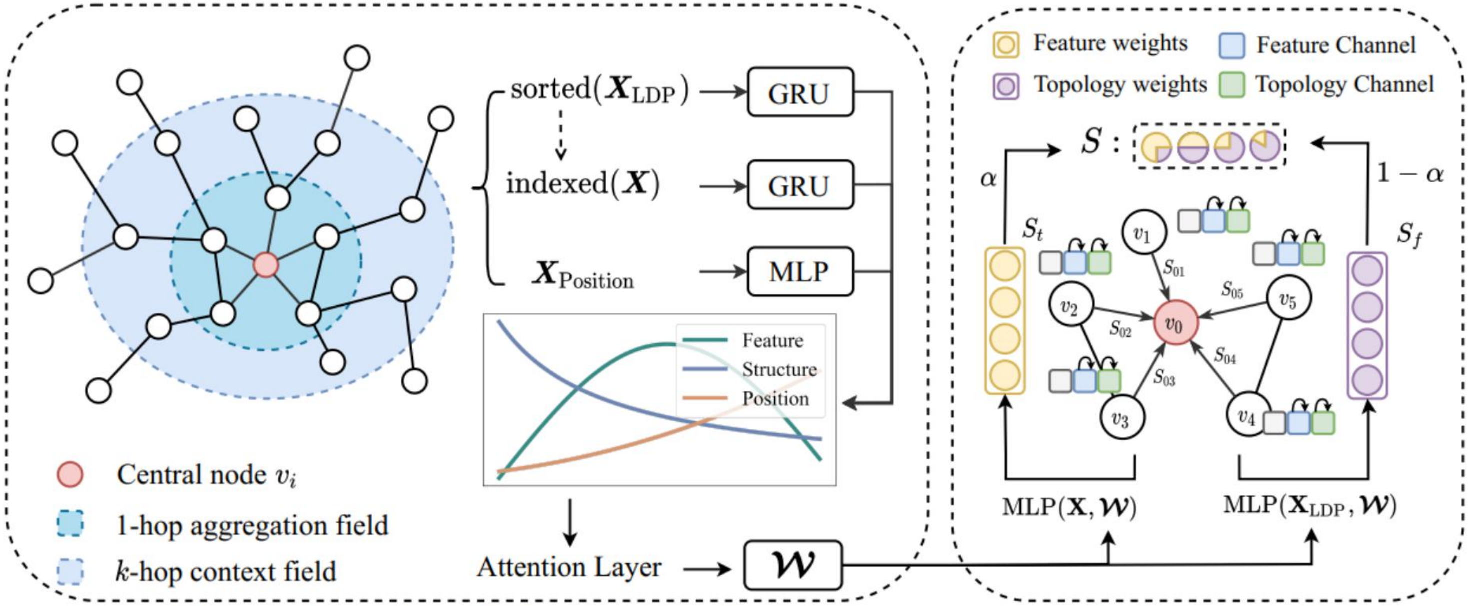
Topological Structure Field.

$$X_{\mathrm{LDP},i} = [d_i, \mathrm{MIN}(\mathrm{DN}_i), \mathrm{MAX}(\mathrm{DN}_i), \mathrm{MEAN}(\mathrm{DN}_i), \mathrm{STD}(\mathrm{DN}_i)],$$

$$\begin{aligned} \boldsymbol{u}^{(t)} &= \sigma\left(\boldsymbol{W}_u\left[\boldsymbol{h}^{(t-1)}, \boldsymbol{x}^{(t)}\right] + \boldsymbol{b}_u\right) \\ \boldsymbol{r}^{(t)} &= \sigma\left(\boldsymbol{W}_r\left[\boldsymbol{h}^{(t-1)}, \boldsymbol{x}^{(t)}\right] + \boldsymbol{b}_r\right) \\ \hat{\boldsymbol{h}}^{(t)} &= \tanh\left(\boldsymbol{W}_h\left[\boldsymbol{r}^{(t)} \odot \boldsymbol{h}^{(t-1)}, \boldsymbol{x}^{(t)}\right] + \boldsymbol{b}_h\right) \\ \boldsymbol{h}^{(t)} &= \left(1 - \boldsymbol{u}^{(t)}\right) \odot \boldsymbol{h}^{(t-1)} + \boldsymbol{u}^{(t)} \odot \hat{\boldsymbol{h}}^{(t)} \end{aligned} \quad (4)$$

$$\mathcal{D}_t = \boldsymbol{h}^{(T)}$$

Method



(a) Modeling Node Local Distribution with Meta-Weight

(b) Adaptive Convolution

Node Feature Field.

$$X_{\text{Feature}} = \text{SORT}\left(\{X_j | v_j \in \mathcal{N}_{i,k}\}\right),$$

$$\mathcal{D}_f$$

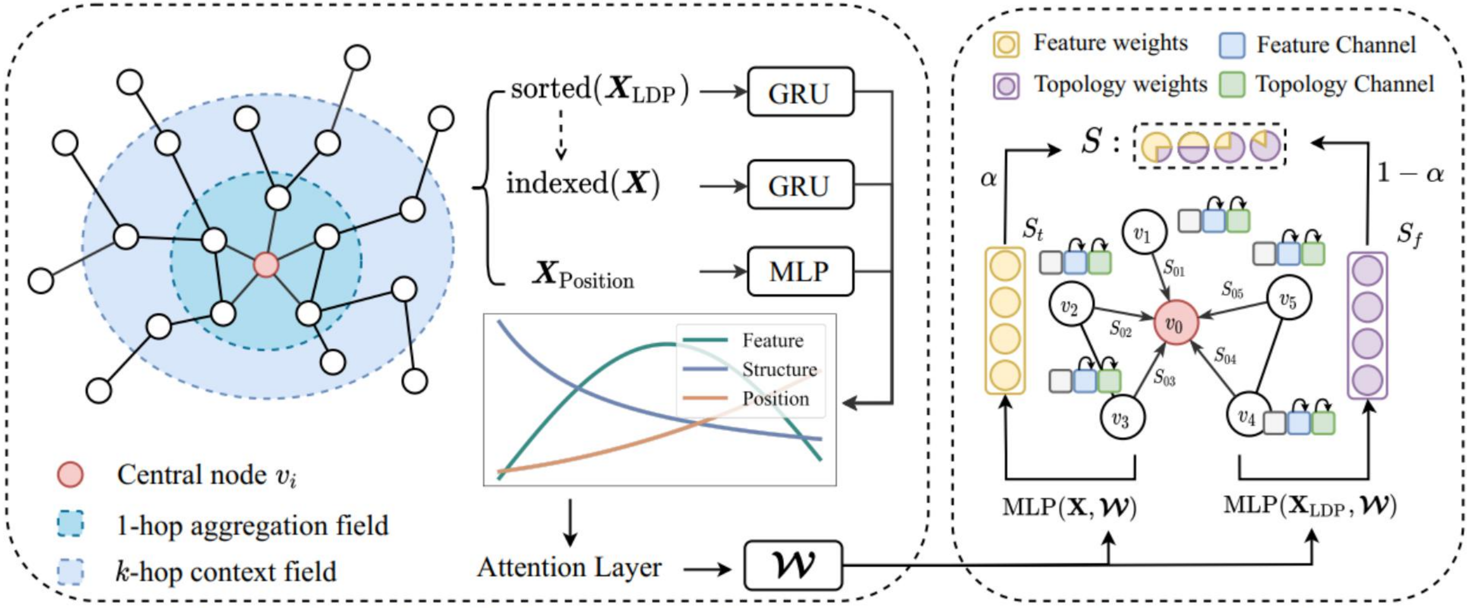
Positional Identity Field.

$$X_{\text{Position},i} = (\phi(v_i, v_1), \phi(v_i, v_2), \cdots, \phi(v_i, v_n)), \quad (5)$$

$\phi(v_i, v_j)$ denotes the shortest path (SPD)

$$\mathcal{D}_p = \Phi\left(X_{\text{Position}}\right)$$

Method



(a) Modeling Node Local Distribution with Meta-Weight (b) Adaptive Convolution

Integration of Three Distributions

$$\omega_t^i = \boldsymbol{q} \cdot \tanh\left(\boldsymbol{W}_a \cdot (\mathcal{D}_{t,i})^T + \boldsymbol{b}\right), \quad (6)$$

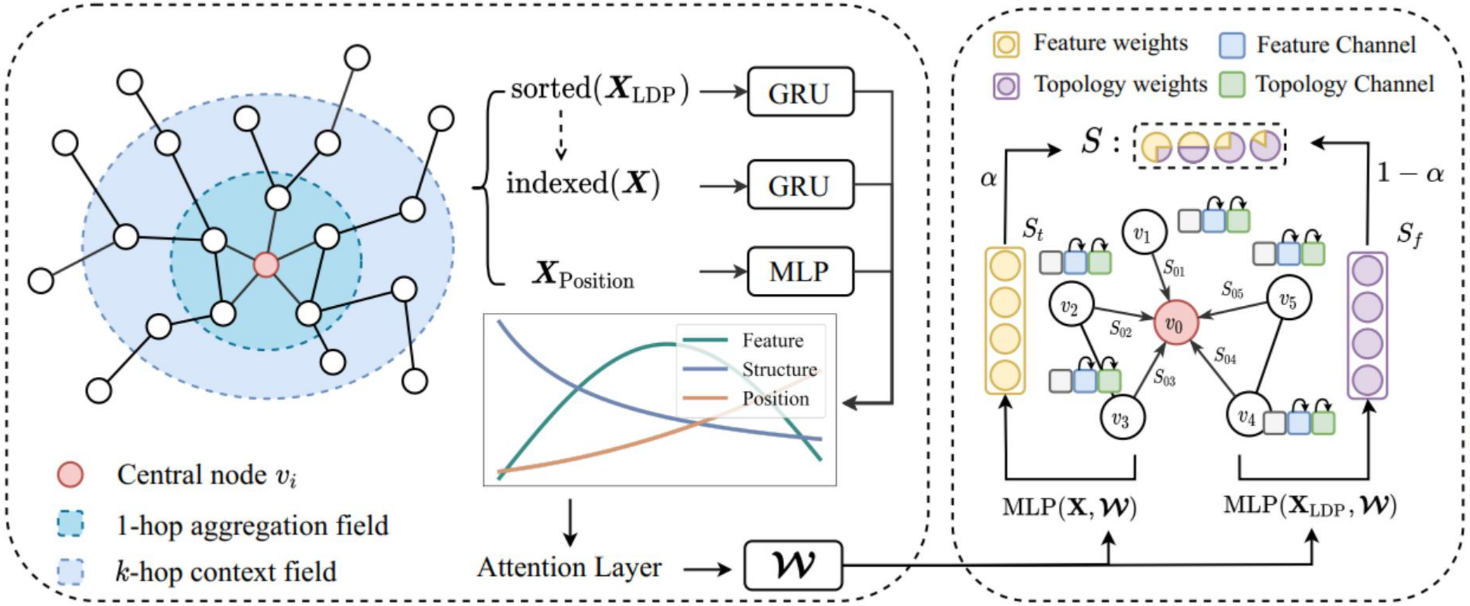
$$a_t^i = \frac{\exp(\omega_t^i)}{\exp(\omega_t^i)+\exp\left(\omega_f^i\right)+\exp\left(\omega_p^i\right)}$$

$$\boldsymbol{a}_t = \left[a_t^i\right], \boldsymbol{a}_f = \left[a_f^i\right], \boldsymbol{a}_p = \left[a_p^i\right] \in \mathbb{R}^{N\times 1}$$

$$\boldsymbol{a}_t = \operatorname{diag}(\boldsymbol{a}_t), \boldsymbol{a}_f = \operatorname{diag}\left(\boldsymbol{a}_f\right) \text{ and } \boldsymbol{a}_p = \operatorname{diag}(\boldsymbol{a}_p)$$

$$\mathcal{W} = \boldsymbol{a}_f \odot \mathcal{D}_f + \boldsymbol{a}_t \odot \mathcal{D}_t + \boldsymbol{a}_p \odot \mathcal{D}_p. \quad (7)$$

Method



(a) Modeling Node Local Distribution with Meta-Weight
(b) Adaptive Convolution

Decouple Topology and Feature in Aggregation.

$$H^{(l+1)} = \sigma\left(\hat{P} H^{(l)} W^{(l)}\right), \hat{P} = A \odot S \tag{8}$$

$$S = \alpha \cdot S_f + (1-\alpha) \cdot S_t \tag{9}$$

$$S_f = \Psi_f\left(\mathcal{W}, X\right), S_t = \Psi_t\left(\mathcal{W}, X_{\mathrm{LDP}}\right), \tag{10}$$

Independent Convolution Channels for Topology and Feature.

$$H^{(l+1)} = \sigma\Big(\left(1-\lambda_1-\lambda_2\right) \hat{P} H^{(l)} + \lambda_1 H_f^{(0)} + \lambda_2 H_t^{(0)}\Big) \cdot \Big(\left(1-\beta\right) \cdot I_n + \beta \cdot W^{(l)}\Big), \tag{11}$$

Experiments

Dataset	$\|\mathcal{V}\|$	$\|\mathcal{E}\|$	$\|\mathcal{Y}\|$	F	h
Cora	2708	10556	7	1433	0.81
Citeseer	3327	9104	6	3703	0.74
Pubmed	19717	88648	3	500	0.8
Texas	183	325	5	1703	0.11
Cornell	183	298	5	1703	0.31
Chameleon	2277	36101	5	1703	0.2
Squirrel	5201	217073	5	2089	0.22

Table 1: The summary of mean and standard deviation of accuracy over all runs. The best results for each dataset is highlighted in gray. "-" stands for Out-Of-Memory.

	Cora	**Citeseer**	**Pubmed**	**Chameleon**	**Squirrel**	**Texas**	**Cornell**
MLP	60.02 ± 0.75	53.36 ± 1.40	63.40 ± 5.03	48.50 ± 2.49	35.38 ± 1.66	75.95 ± 5.06	77.13 ± 5.32
GCN	80.50 ± 0.50	70.80 ± 0.50	79.00 ± 0.30	38.22 ± 2.67	27.12 ± 1.45	58.05 ± 4.81	56.87 ± 5.29
GAT	83.00 ± 0.70	72.50 ± 0.70	79.00 ± 0.30	43.07 ± 2.31	31.70 ± 1.85	57.38 ± 4.95	54.95 ± 5.63
GPR-GNN	83.69 ± 0.47	71.51 ± 0.29	79.77 ± 0.27	49.56 ± 1.71	37.21 ± 1.15	80.81 ± 2.55	78.38 ± 4.01
CPGNN-MLP-1	79.50 ± 0.38	71.76 ± 0.22	77.45 ± 0.24	49.25 ± 2.83	33.17 ± 1.87	80.00 ± 4.22	80.13 ± 6.47
CPGNN-MLP-2	78.21 ± 0.93	71.99 ± 0.39	78.26 ± 0.33	51.24 ± 2.43	28.86 ± 1.78	79.86 ± 4.64	79.05 ± 7.78
CPGNN-Cheby-1	81.13 ± 0.21	69.72 ± 0.59	77.79 ± 1.06	48.29 ± 2.02	36.17 ± 2.87	76.89 ± 4.95	75.00 ± 7.64
CPGNN-Cheby-2	77.68 ± 1.55	69.92 ± 0.46	78.81 ± 0.28	50.95 ± 2.46	31.29 ± 1.26	76.89 ± 5.83	75.27 ± 7.80
AM-GCN	81.70 ± 0.71	71.72 ± 0.55	-	56.70 ± 3.44	-	74.41 ± 4.50	74.11 ± 5.53
H2GCN	81.85 ± 0.38	70.64 ± 0.65	79.78 ± 0.43	59.39 ± 1.58	37.90 ± 2.02	75.13 ± 4.95	78.38 ± 6.62
MWGNN	83.30 ± 0.62	72.90 ± 0.47	82.30 ± 0.64	79.54 ± 1.28	75.41 ± 1.83	81.37 ± 4.27	79.24 ± 5.23

Experiments

Table 2: Ablation Study: Accuracy of MWGNN and its variants on three synthetic combined graph.

	C.Heter	C.Mixed	C.Homo
MWGNN	56.28	76.38	95.98
w/o D	47.24	69.84	94.73
w/o D_f	51.75	74.38	96.48
w/o D_t	50.76	71.32	94.98
w/o D_p	52.26	73.78	95.21
w/o Indep. Channels	53.77	73.87	86.73

Experiments

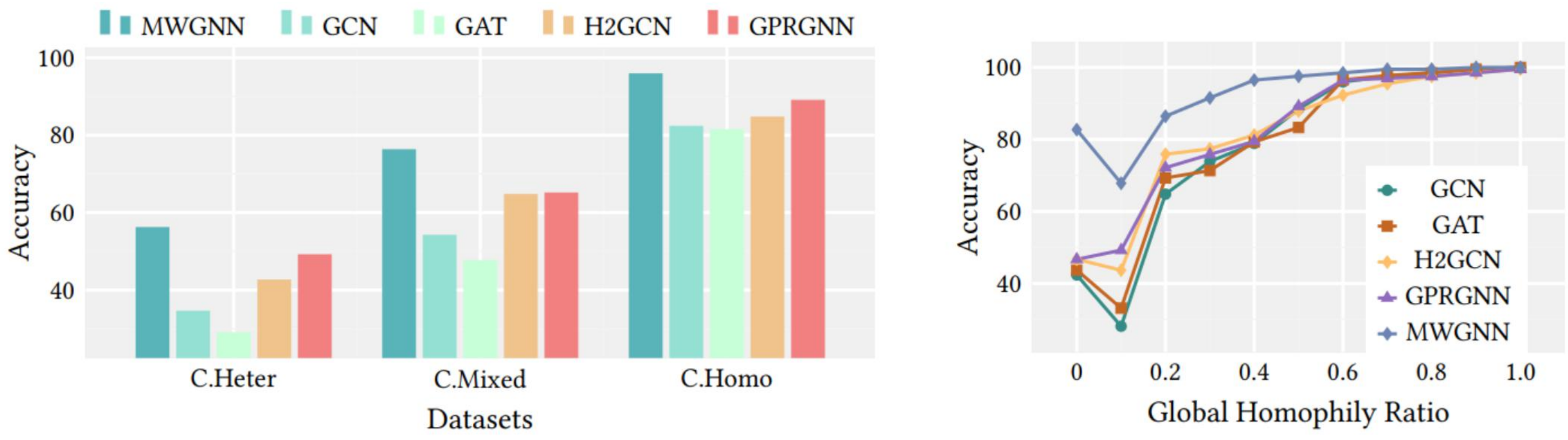


Figure 3: MWGNN and other baselines on synthetic datasets.

Experiments

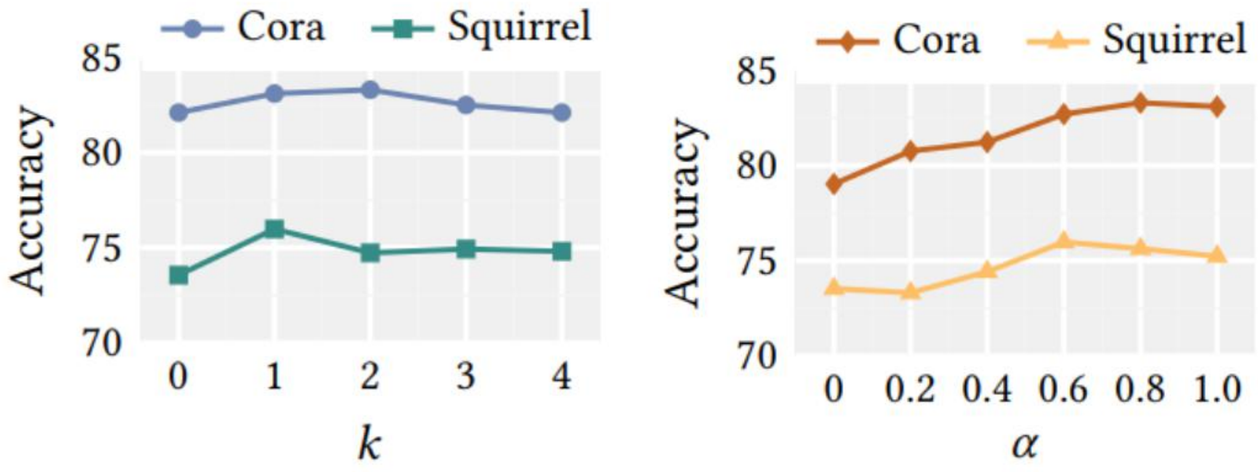


Figure 4: Parameter analysis over Cora and Squirrel on hop number k and combine alpha α.

Thanks