Adaptive Multi-Modalities Fusion in Sequential Recommendation

Hengchang Hu[*†]
hengchang.hu@u.nus.edu
National University of Singapore
Singapore

Wei Guo
guowei67@huawei.com
Huawei Noah's Ark Lab
Singapore

Yong Liu
liu.yong6@huawei.com
Huawei Noah's Ark Lab
Singapore

Min-Yen Kan
kanmy@comp.nus.edu.sg
National University of Singapore
Singapore

code: https://github.com/HoldenHu/MMSR.

CIKM 2023

Reported by Minqin Li

Introduction

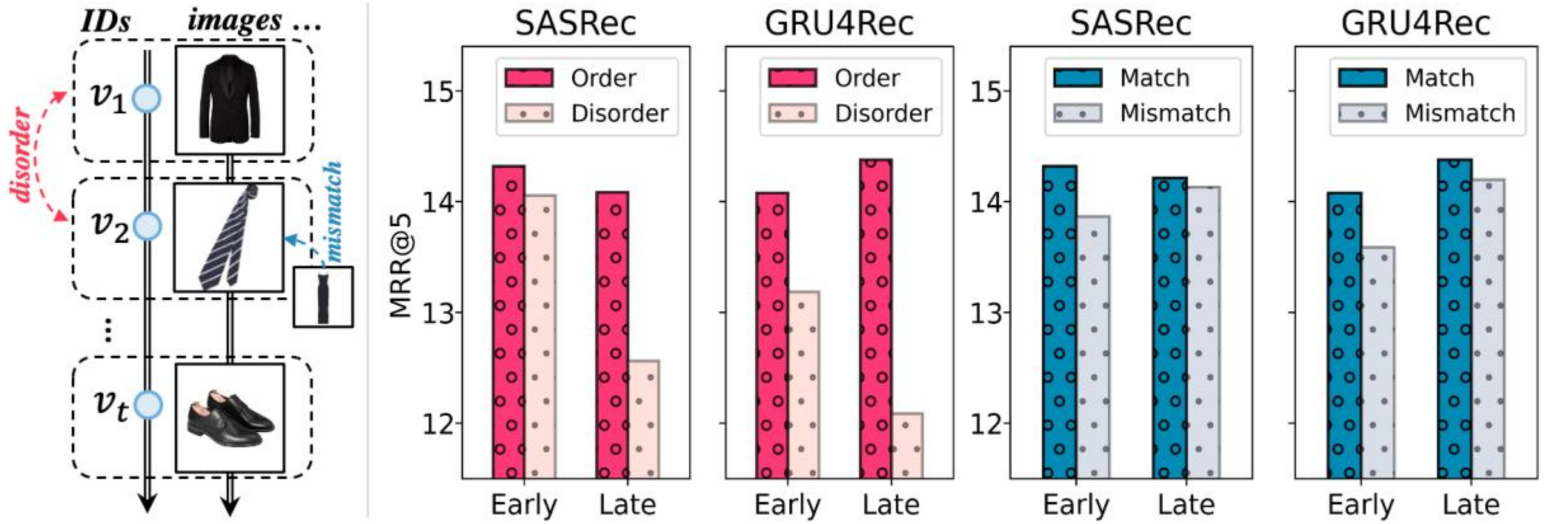

Figure 1: Case study on the Amazon-Fashion dataset. Here, *Order/Match* refers to the original modality sequence, while *disordered* refers to a shuffled item order sequence, and *mis-matched* refers to a condition with displaced modalities.

Early fusion is less sensitive to the interactions between intra-channel features.

Late fusion is less sensitive to the interactions among different channels of features.

Method

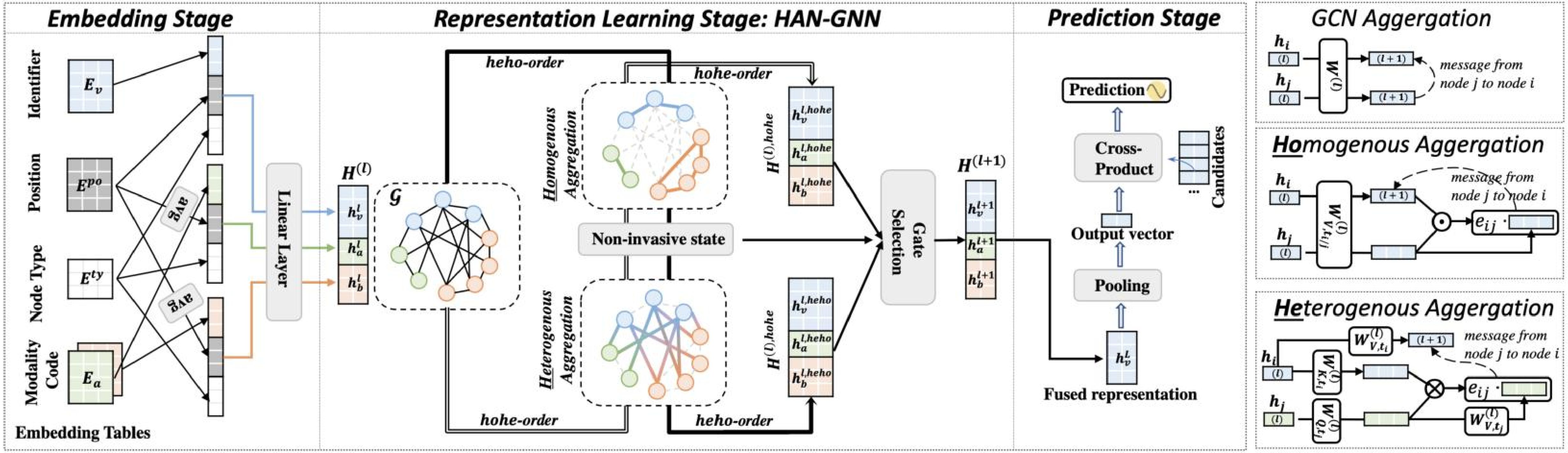

Figure 2: Overall framework of MMSR (left), and the applied aggregation modules (right). Distinct node types are represented by different colors.

Method

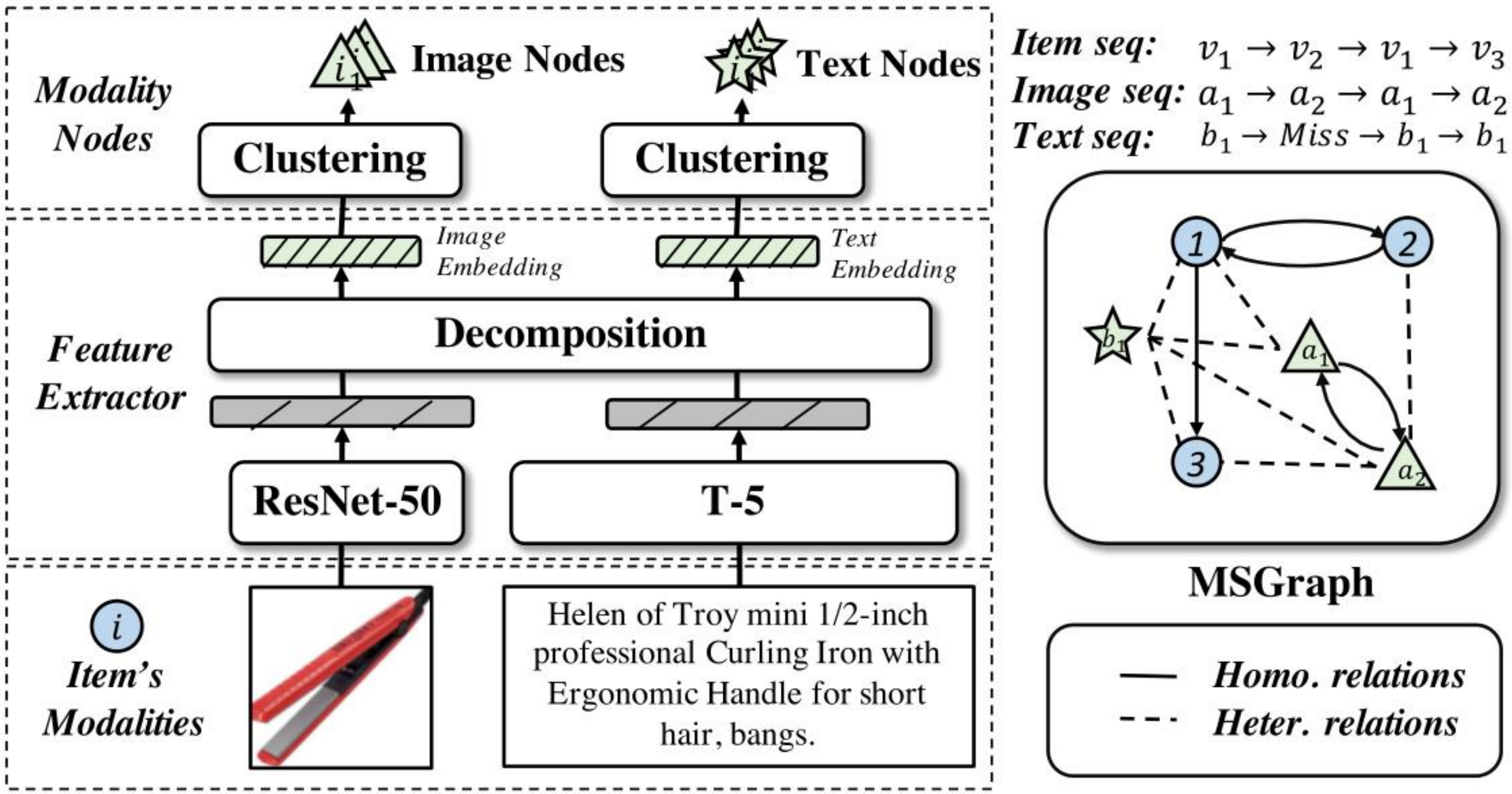

Figure 3: Modality-enriched graph construction.

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_{v_1}, & \mathbf{e}_{v_2}, & \cdots, & \mathbf{e}_{v_m} \\ \mathbf{e}_{a_1}, & \mathbf{e}_{a_2}, & \cdots, & \mathbf{e}_{a_m} \\ \mathbf{e}_{b_1}, & \mathbf{e}_{b_2}, & \cdots, & \mathbf{e}_{b_m} \end{bmatrix} \tag{1}$$

$$\mathbf{P} = f(\mathbf{E}) \tag{2}$$

Early fusion:

$$\mathbf{E}_{i,:} = \sigma(\mathbf{W}(cat[\mathbf{E}_{i,1}; \mathbf{E}_{i,2}; \mathbf{E}_{i,3}])) \tag{3}$$

$$\mathbf{P} = \mathcal{M}([\mathbf{E}_{1,:}, \mathbf{E}_{2,:}, ..., \mathbf{E}_{m,:}]) \tag{4}$$

Late fusion:

$$\mathbf{E}_{:,j} = \mathcal{M}([\mathbf{E}_{1,j}, \mathbf{E}_{2,j}, ..., \mathbf{E}_{m,j}]) \tag{5}$$

$$\mathbf{P} = \sigma(\mathbf{W}(cat[\mathbf{E}_{:,1}; \mathbf{E}_{:,2}; \mathbf{E}_{:,3}])) \tag{6}$$

$$\hat{y} = < \mathbf{P}, \mathbf{e}_v^\top > \tag{7}$$

Method

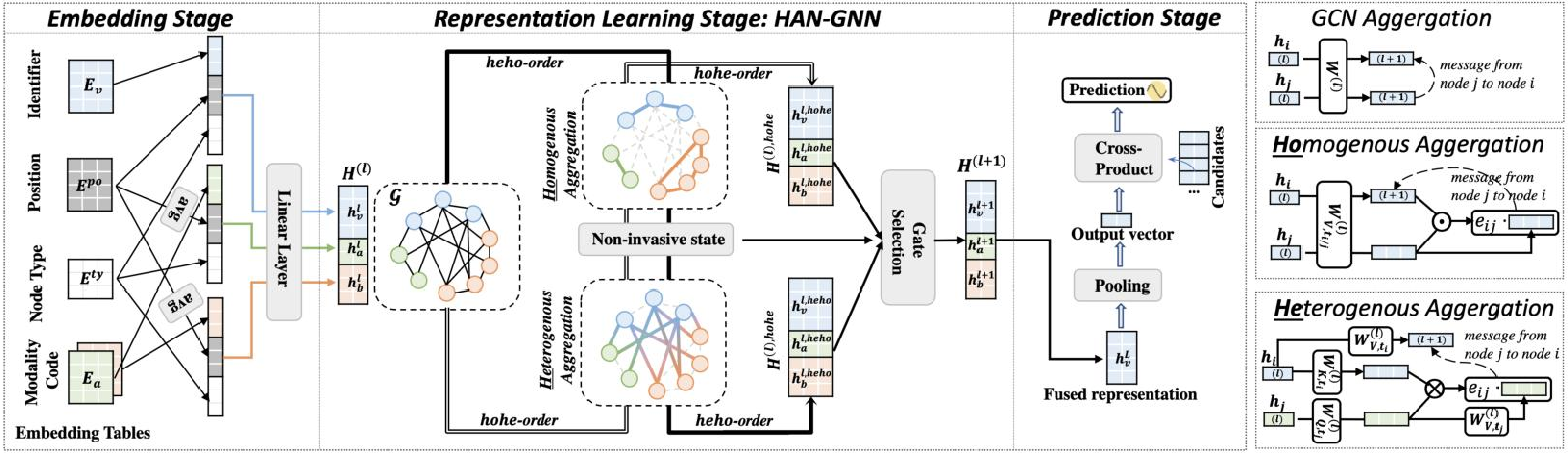

Figure 2: Overall framework of MMSR (left), and the applied aggregation modules (right). Distinct node types are represented by different colors.

$$h_i^{(l+1)} = \sigma\left(\sum\nolimits_{j\in N_i} d(i,j) W^{(l)} h_j^{(l)}\right) \tag{8}$$

$$h_i^{(l+1)} = \sum\nolimits_{j\in N_i} \alpha_{ij}^{(l)} h_j^{(l)} \tag{9}$$

$$e_{ij}^{(l)} = a^T [W^{(l)} h_i^{(l)}; W^{(l)} h_j^{(l)}] \tag{10}$$

$$\alpha_{ij}^{(l)} = sft(e_{ij}^{(l)} | N_i) = \frac{\exp(\mathrm{LeakyReLU}(e_{ij}^{(l)}))}{\sum_{k\in N_i} \exp(\mathrm{LeakyReLU}(e_{ik}^{(l)}))} \tag{11}$$

Method

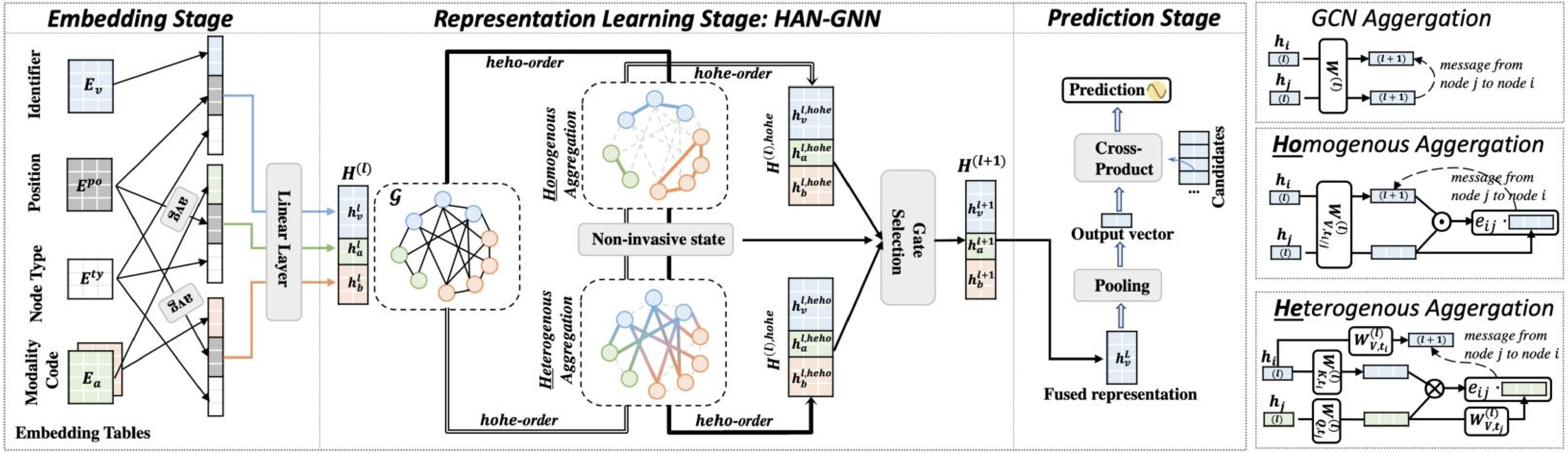

Figure 2: Overall framework of MMSR (left), and the applied aggregation modules (right). Distinct node types are represented by different colors.

$$e_{ij}^{(l),ho} = a_r(W_{V,t_i}^{(l)} h_i^{(l)} \odot W_{V,t_j}^{(l)} h_j^{(l)}) \quad (12)$$

$$e_{ij}^{(l),he} = (W_{Q,t_j}^{(l)} h_j^{(l)})(W_{K,t_i}^{(l)} h_i^{(l)})^\top \quad (13)$$

$$h_i^{(l+1),*} = \sum\nolimits_{j \in N_i} sft(e_{ij}^{(l),*} | N_i)(W_{V,t_j}^{(l)} h_j^{(l)}) \quad (14)$$

$$h_i^{(l+1)} = Linear([h_i^{(l+1),ho}; h_i^{(l+1),he}]) \quad (15)$$

Method

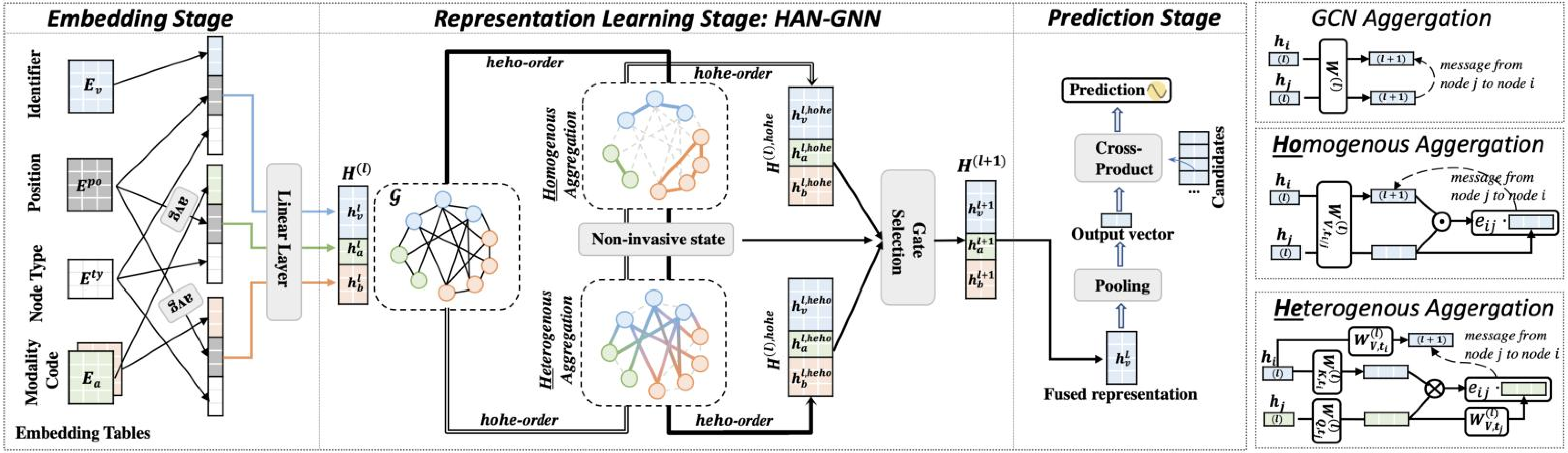

Figure 2: Overall framework of MMSR (left), and the applied aggregation modules (right). Distinct node types are represented by different colors.

$$\beta = MLP([h_i^{(l+1),hohe}; h_i^{(l+1),heho}]) \quad (16)$$

$$h_i^{(l+1)} = \beta_0 \times h_i^{(l+1),hohe} + \beta_1 \times h_i^{(l+1),heho} \quad (17)$$

Experiments

	Beauty	Clothing	Sports	Toys	Kitchen	Phone
User #	22,363	39,387	35,598	19,412	27,879	66,519
Item #	12,101	23,033	18,357	11,924	10,429	28,237
Inter. #	198,502	278,677	296,337	167,597	194,439	551,682
Avg Len. #	8.88	7.12	8.46	8.79	7.19	8.35
Sparsity	99.93%	99.97%	99.95%	99.93%	99.93%	99.97%

Table 1: Dataset Statistics after preprocessing.

Experiments

	Metric	GRU4Rec	SASRec	SR-GNN	MMGCN	MGAT	BM3	GRU4Rec^F	SASRec^F	NOVA	DIF-SR	Trans2D	MMSR
Beauty	HR@5	5.6420	6.1900	4.1483	2.6534	4.0870	4.8713	3.7682	6.4021	4.2219	6.5789	6.0191	**7.1563***
	MRR@5	3.1110	3.2165	2.2123	1.2534	2.0297	2.3349	2.0793	3.7990	2.1785	4.0735	3.4387	**4.4429***
	HR@20	12.7217	14.0681	10.2351	7.0443	9.1126	10.2640	9.4868	14.0269	10.7978	14.0137	13.2214	**14.1470***
	MRR@20	3.7714	3.9668	2.7911	1.5263	2.6714	3.1945	2.6006	4.5073	2.8160	4.7983	3.9460	**5.0433***
Clothing	HR@5	1.3340	1.5885	0.8547	0.5231	0.9613	1.2851	0.9501	1.8430	1.2937	1.5524	1.3929	**1.8684***
	MRR@5	0.6765	0.7820	0.4555	0.2128	0.5470	0.5460	0.5212	0.9470	0.6503	0.7961	0.6682	**1.1365***
	HR@20	3.8111	3.9574	2.7528	1.7847	2.7363	3.5072	2.8610	4.2048	3.4866	4.0571	4.0683	**4.4136***
	MRR@20	0.9418	1.0339	0.6251	0.4359	0.7548	0.9045	0.6955	1.2814	0.8783	1.0530	1.0391	**1.3344***
Sport	HR@5	2.4388	2.9549	2.0742	1.2020	2.0418	2.3096	1.8929	3.1063	2.1539	2.5145	2.7168	**3.2657***
	MRR@5	1.2696	1.5858	1.0790	0.5688	0.8762	0.9963	0.9786	1.6997	1.1271	1.3469	1.4235	**1.9846***
	HR@20	6.6430	7.2208	5.4376	3.6492	5.2197	5.3184	5.4834	7.3683	5.8062	7.0774	6.9453	**7.7466***
	MRR@20	1.6947	2.0357	1.4349	0.8645	1.3002	1.5245	1.3274	2.1427	1.5648	1.9214	1.7058	**2.2826***
Toys	HR@5	3.8663	5.0902	2.7329	1.7592	2.3746	3.9084	2.1974	5.2328	3.7899	5.2363	4.1908	**6.1159***
	MRR@5	2.0022	2.7536	1.4878	0.7869	1.1369	2.0352	1.1576	3.0801	1.9641	3.1944	2.2370	**3.8987***
	HR@20	10.0727	11.8668	6.7452	4.5497	5.9223	8.7071	6.0638	11.7485	9.0609	12.0284	10.5082	**12.1192***
	MRR@20	2.7267	3.4228	1.8655	1.1256	1.5314	2.5623	1.5230	3.6812	2.4502	3.8777	2.9298	**4.3551***
Kitchen	HR@5	1.1759	1.8012	1.1024	0.6671	1.2225	1.4399	1.1323	1.9077	1.2558	1.5828	1.3463	**2.2145***
	MRR@5	0.5824	0.9729	0.5877	0.3154	0.4882	0.7012	0.5586	1.1268	0.6279	0.8499	0.7413	**1.4238***
	HR@20	3.5640	4.2021	3.3255	2.2404	3.5206	3.4157	3.5449	4.3187	3.5332	4.2766	3.8158	**4.4535***
	MRR@20	0.8277	1.2043	0.8507	0.5210	0.6898	0.8832	0.7817	1.3862	0.8349	1.1041	0.8682	**1.6086***
Phone	HR@5	5.6626	6.4435	5.3128	3.2823	4.4046	4.9338	4.1188	6.6908	5.3581	6.0666	6.0646	**6.9550***
	MRR@5	2.8765	3.4998	2.7221	1.4397	1.8735	2.3515	2.0211	3.6643	2.7899	3.2383	3.0125	**3.9911***
	HR@20	13.4539	14.1525	12.1363	8.3255	10.9956	11.0081	11.3945	14.6771	12.3232	14.6781	13.8446	**14.9509***
	MRR@20	3.7002	4.3182	3.4807	2.0647	3.0360	3.2278	3.0653	4.5001	3.5063	4.2540	3.8798	**4.5747***

Table 2: Overall Performance (%). Bold ones indicate the best performances, while underlined ones indicate the best among baselines. $*$ indicates a statistically significant level p-value<0.05 comparing MMSR with the best baseline.

Experiments

Model	Beauty		Clothing		Sport	
	HR@5	MRR@5	HR@5	MRR@5	HR@5	MRR@5
GCN	5.6348	3.163	1.2340	0.6465	2.3177	1.1424
GraphSAGE	5.5773	3.1283	1.3801	0.8552	2.2496	1.3473
GAT	5.7116	3.1941	1.4092	0.8332	2.3452	1.3825
Graphormer	5.9267	3.3029	1.4573	0.9029	2.3069	1.3756
RGAT	6.8157	3.9783	1.7352	1.0873	2.8609	1.7133
HGNN	6.9701	4.1276	1.7721	1.1084	2.9682	1.7776
HGAT	7.0671	4.2494	1.8448	1.1417	3.0458	1.8501
HAN-GNN	7.1386	4.6244	2.0402	1.2642	3.3255	1.9916

Table 3: Graph Aggregator Comparison.

Experiments

Model	Beauty		Clothing		Sport	
	HR@5	MRR@5	HR@5	MRR@5	HR@5	MRR@5
HAN-GNN	7.1386	4.6244	2.0402	1.2642	3.3255	1.9916
Synchronous	6.8912	4.4515	1.7857	1.0681	3.0924	1.7849
$NI(hohe)$	6.8900	4.4616	1.9999	1.2357	3.0616	1.8792
$NI(heho)$	6.8897	4.5528	1.3932	0.7655	3.0087	1.7051
hohe	6.8971	4.4245	1.9575	1.2169	3.0565	1.8793
heho	6.5406	4.3117	1.1495	0.6398	2.8871	1.6654
ho	6.6702	4.1004	1.6012	0.9069	3.0306	1.7412
he	6.9354	4.446	1.9957	1.2236	3.0047	1.8648
w/o $\mathbf{e}^{po}$	6.9664	4.5653	2.0665	1.2581	3.1547	1.8968
w/o $\mathbf{e}^{ty}$	6.9390	4.5074	2.0370	1.2593	3.2112	1.9854

Table 4: Ablation analysis, evaluated with (HR, MRR)@5. The relation ablation is based on a GCN aggregator.

Experiments

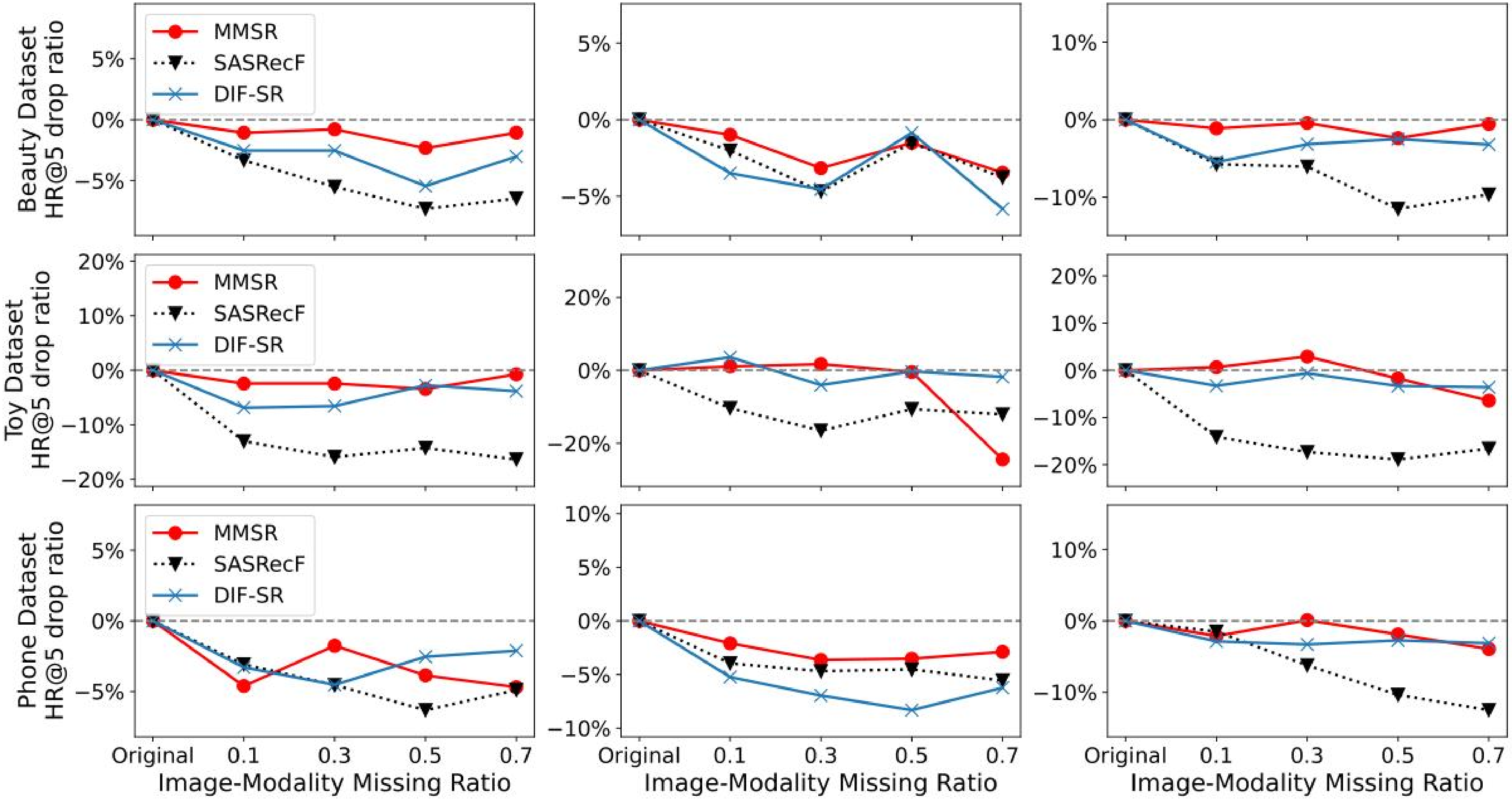

Figure 4: MMSR robustness against missing modalities.

Experiments

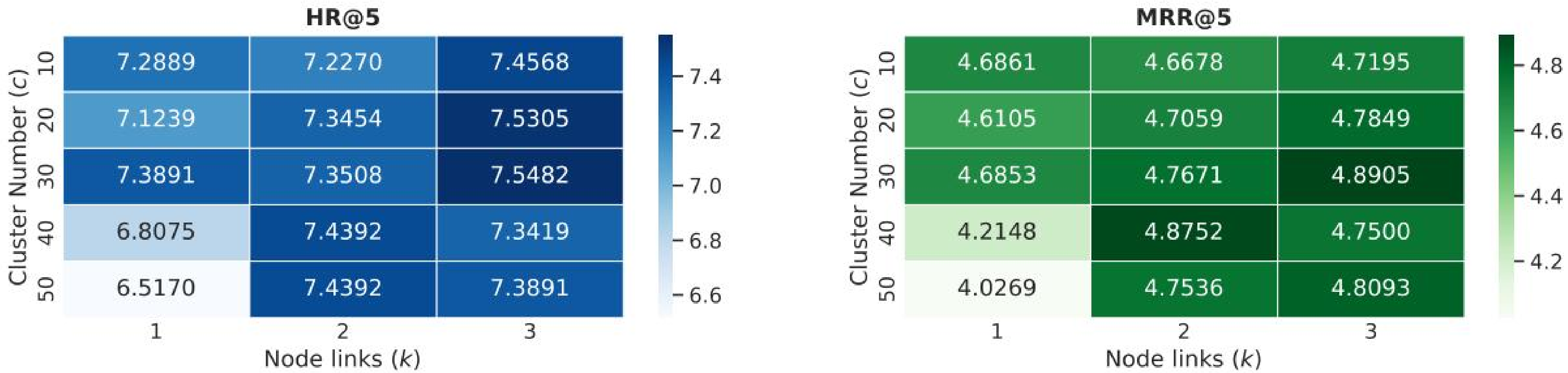

Figure 5: The performance comparison with different MS-Graph construction parameters on the Beauty dataset.

Thanks