Adaptive Graph Contrastive Learning for Recommendation

Yangqin Jiang
University of Hong Kong
Hong Kong, China
mrjiangyq99@gmail.com

Chao Huang*
University of Hong Kong
Hong Kong, China
chaohuang75@gmail.com

Lianghao Xia
University of Hong Kong
Hong Kong, China
aka_xia@foxmail.com

https://github.com/HKUDS/AdaGCL.

（KDD-2023)

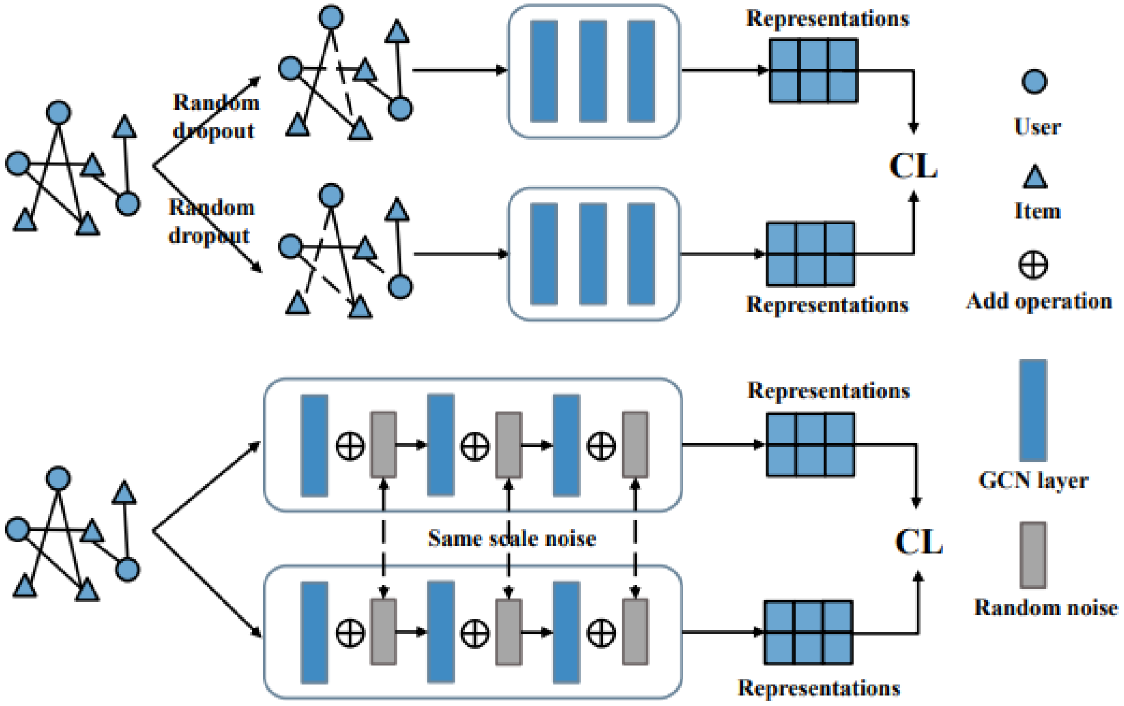
Reported by Ke Gan

1. Introduction
2. Approach
3. Experiments

gesis
Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center
L3S

Introduction



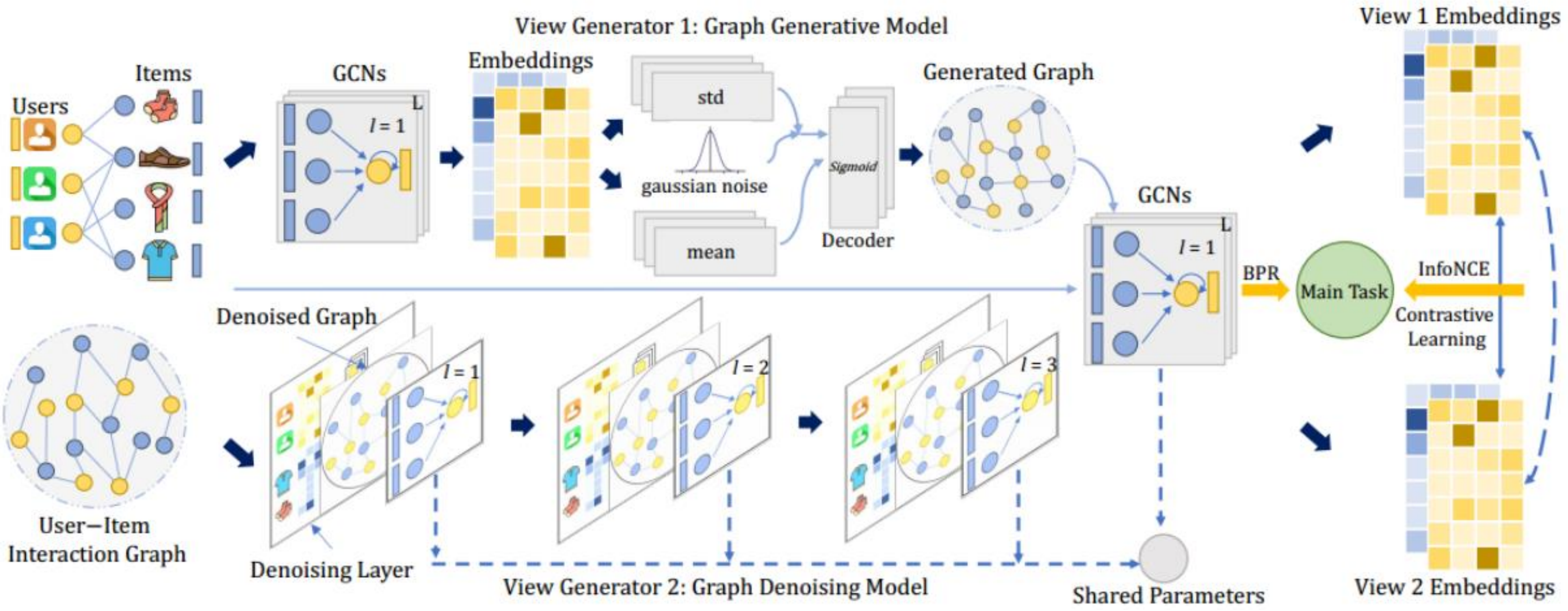
data noise, such as users clicking on irrelevant products

the sparsity and skewed distribution of recommendation data can negatively impact effective user-item interaction modeling.

self-supervised learning creating contrastive views through probability-based random masking or adding noise.

may keep some noisy interactions or drop important training signals

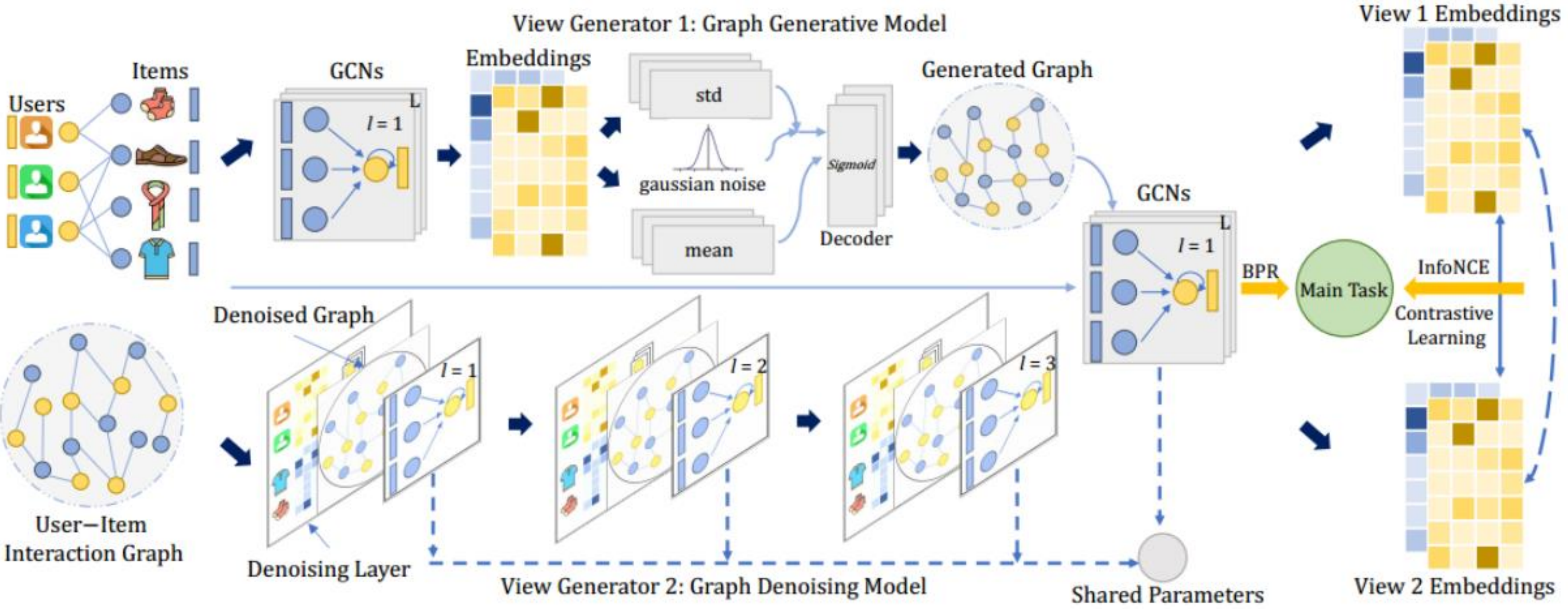
Approach



$\mathcal{U} = \{u_1, \cdots, u_i, \cdots, u_I\}$ interaction matrix $\mathcal{A} \in \mathbb{R}^{I \times \mathcal{J}}$

$\mathcal{V} = \{v_1, \cdots, v_j, \cdots, v_J\}$ embedding matrices $\mathbf{E}^{(u)} \in \mathbb{R}^{I \times d}$ and $^{v)} \in \mathbb{R}^{J \times d}$

Approach



$$\mathbf{z}_i^{(u)} = \bar{\mathcal{A}}_{i,*}\cdot \mathbf{E}^{(v)}, \quad \mathbf{z}_j^{(v)} = \bar{\mathcal{A}}_{*,j}\cdot \mathbf{E}^{(u)}, \tag{1}$$

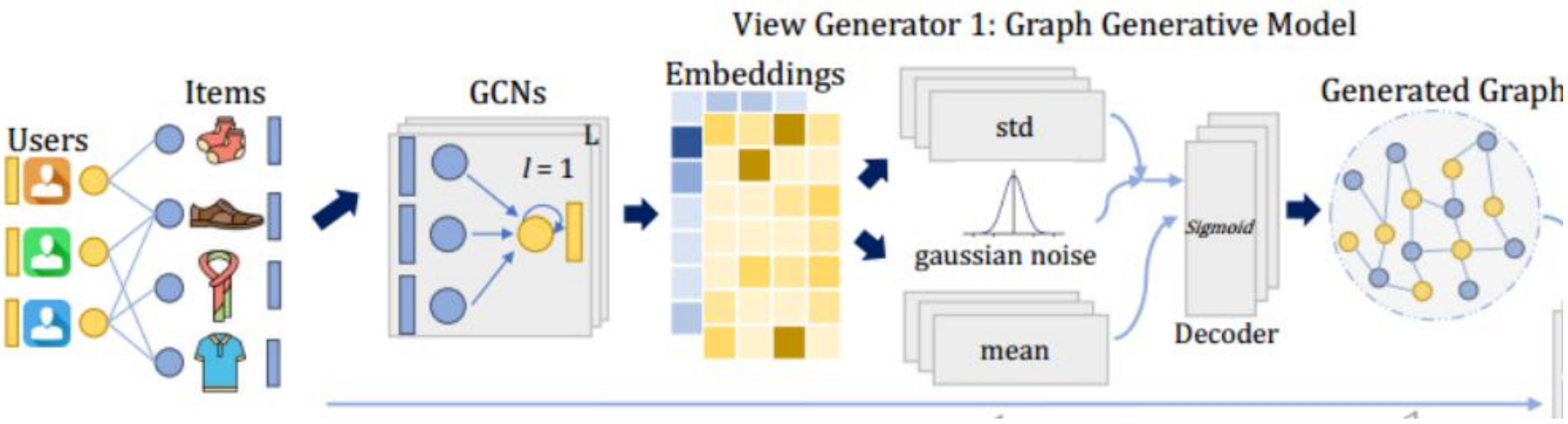
$$\bar{\mathcal{A}} = \mathbf{D}_{(u)}^{-1/2}\cdot \mathcal{A}\cdot \mathbf{D}_{(v)}^{-1/2}, \quad \bar{\mathcal{A}}_{i,j} = \frac{\mathcal{A}_{i,j}}{\sqrt{|\mathcal{N}_i|\cdot|\mathcal{N}_j|}}, \tag{2}$$

$$\mathbf{e}_{i,l}^{(u)} = \mathbf{z}_{i,l}^{(u)} + \mathbf{e}_{i,l-1}^{(u)}, \quad \mathbf{e}_{j,l}^{(v)} = \mathbf{z}_{j,l}^{(v)} + \mathbf{e}_{j,l-1}^{(v)}. \tag{3}$$

$$\mathbf{e}_i^{(u)} = \sum_{l=0}^{L}\mathbf{e}_{i,l}^{(u)}, \quad \mathbf{e}_j^{(v)} = \sum_{l=0}^{L}\mathbf{e}_{j,l}^{(v)}, \quad \hat{y}_{i,j} = \mathbf{e}_i^{(u)\top}\mathbf{e}_j^{(v)}. \tag{4}$$

$$\mathcal{L}_{ssl}^{user} = \sum_{u_i\in\mathcal{U}} -\log\frac{\exp(s(\mathbf{e}_i', \mathbf{e}_i'')/\tau)}{\sum_{u_{i'}\in\mathcal{U}}\exp(s(\mathbf{e}_i', \mathbf{e}_{i'}'')/\tau)}, \tag{5}$$

Approach



Two MLPs are utilized to derive the $\boxed{\text{mean value}}$ and the $\boxed{\text{standard deviation}}$

another MLP as the decoder, $p(A_{ij} = 1 | z_i, z_j) = \sigma(z_i^T z_j)$

$$\mu = GCN_\mu(X, A)$$

$$log\sigma = GCN_\sigma(X, A)$$

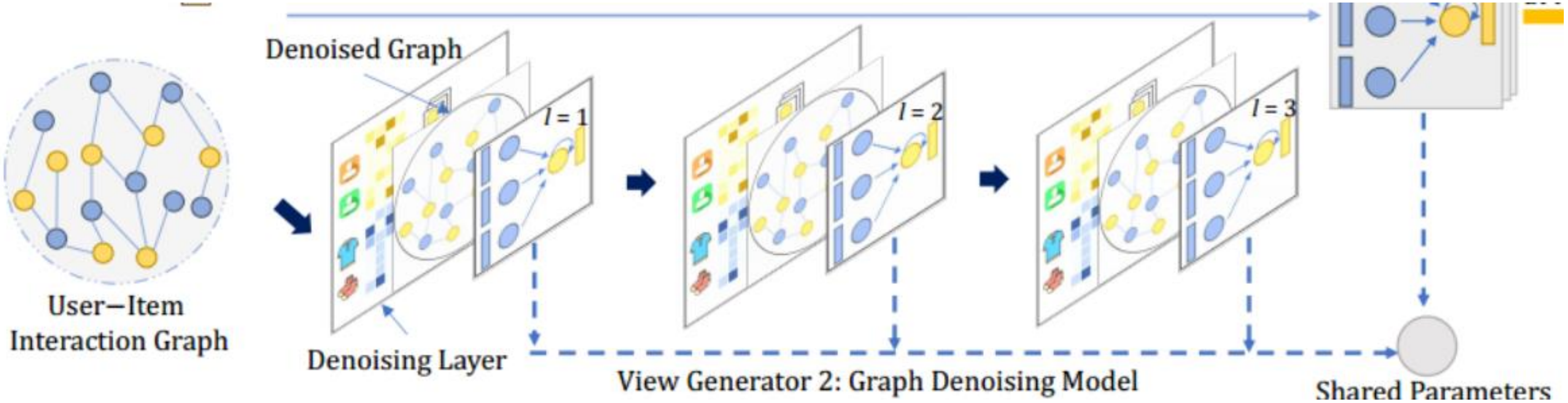
$$GCN(X, A) = \tilde{A}ReLU(\tilde{A}XW_0)W_1$$

$$\mathcal{L}_{gen} = \mathcal{L}_{kl} + \mathcal{L}_{dis}, \tag{6}$$

$$L = E_{q(Z|X,A))}[logp(A|Z)] - KL[q(Z|X,A)||p($$

$$\mathcal{L}_{gen} = \mathcal{L}_{kl} + \mathcal{L}_{dis} + \mathcal{L}_{bpr}^{gen} + \lambda_2||\Theta||_F^2, \tag{10}$$

Approach



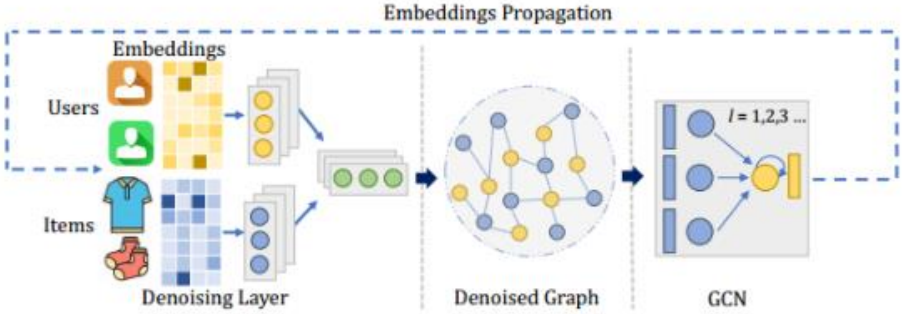


Figure 2: Workflow of the graph denoising model.

binary matrix $\mathbf{M}^l \in 0,1^{|\mathcal{V}|\times|\mathcal{V}|}$,

(0 indicates a noisy edge)

$$\mathbf{A}^l = \mathbf{A} \odot \mathbf{M}^l$$

$$\sum_{l=1}^{L} ||\mathbf{M}^l||_0 = \sum_{l=1}^{L} \sum_{(u,v)\in\varepsilon} \mathbb{I}[m_{i,j}^l \neq 0], \quad (7)$$

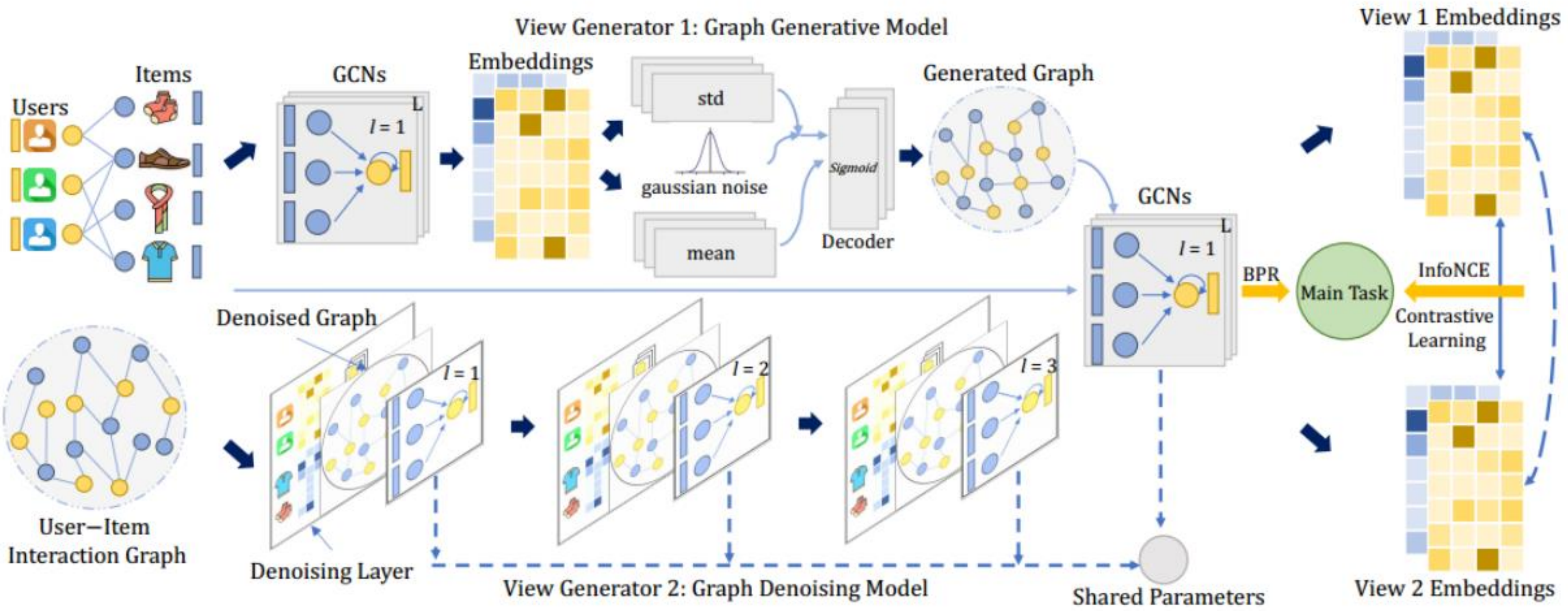
we adopt the reparameterization trick and relax the binary entries $m_{i,j}^l$ from being drawn from a Bernoulli distribution to a deterministic function g of parameters $\alpha_{i,j}^l \in \mathbb{R}$ and an independent random variable ε^l. That is $m_{i,j}^l = g(\alpha_{i,j}^l, \varepsilon^l)$.

$$\alpha_{i,j}^l = f_{\theta^l}^l(\mathbf{e}_i^l, \mathbf{e}_j^l),$$

$$\mathcal{L}_c = \sum_{l=1}^{L} \sum_{(u_i,v_j)\in\varepsilon} (1 - \mathbb{P}_{\sigma(s_{i,j}^l)}(0|\theta^l)), \quad (8)$$

$$\mathcal{L}_{den} = \mathcal{L}_c + \mathcal{L}_{bpr}^{den} + \lambda_2||\Theta||_F^2. \quad (11)$$

Approach



$$\mathcal{L}_{bpr} = \sum_{(u,i,j)\in O} -\log\sigma(\hat{y}_{ui} - \hat{y}_{uj}), \tag{9}$$

$$\mathcal{L}_{upper} = \mathcal{L}_{bpr} + \lambda_1 \mathcal{L}_{ssl} + \lambda_2 ||\Theta||_F^2, \tag{12}$$

$$\mathcal{L}_{lower} = \mathcal{L}_{gen} + \mathcal{L}_{den}. \tag{13}$$

Experiment

Table 2: Performance comparison on Last.FM, Yelp, BeerAdvocate datasets in terms of *Recall* and *NDCG*.

Dataset	Metric	BiasMF	NCF	AutoR	PinSage	STGCN	GCMC	NGCF	GCCF	LightGCN	SLRec	NCL	SGL	HCCF	SHT	DirectAU	Ours	p-val.
Last.FM	Recall@20	0.1879	0.1130	0.1518	0.1690	0.2067	0.2218	0.2081	0.2222	0.2349	0.1957	0.2353	0.2427	0.2410	0.2420	0.2422	**0.2603**	$2.1e^{-5}$
	NDCG@20	0.1362	0.0795	0.1114	0.1228	0.1528	0.1558	0.1474	0.1642	0.1704	0.1442	0.1715	0.1761	0.1773	0.1770	0.1727	**0.1911**	$9.5e^{-5}$
	Recall@40	0.2660	0.1693	0.2174	0.2402	0.2940	0.3149	0.2944	0.3083	0.3220	0.2792	0.3252	0.3405	0.3232	0.3235	0.3356	**0.3531**	$6.9e^{-5}$
	NDCG@40	0.1653	0.0952	0.1336	0.1472	0.1821	0.1897	0.1829	0.1931	0.2022	0.1737	0.2033	0.2104	0.2051	0.2055	0.2042	**0.2204**	$5.6e^{-4}$
Yelp	Recall@20	0.0532	0.0304	0.0491	0.0510	0.0562	0.0584	0.0681	0.0742	0.0761	0.0665	0.0806	0.0803	0.0789	0.0794	0.0818	**0.0873**	$1.5e^{-6}$
	NDCG@20	0.0264	0.0143	0.0222	0.0245	0.0282	0.0280	0.0336	0.0365	0.0373	0.0327	0.0402	0.0398	0.0391	0.0395	0.0424	**0.0439**	$1.8e^{-8}$
	Recall@40	0.0802	0.0487	0.0692	0.0743	0.0856	0.0891	0.1019	0.1151	0.1175	0.1032	0.1230	0.1226	0.1210	0.1217	0.1226	**0.1315**	$3.2e^{-6}$
	NDCG@40	0.0321	0.0187	0.0268	0.0315	0.0355	0.0360	0.0419	0.0466	0.0474	0.0418	0.0505	0.0502	0.0492	0.0497	0.0524	**0.0548**	$2.7e^{-7}$
BeerAdvocate	Recall@20	0.0996	0.0729	0.0816	0.0930	0.1003	0.1082	0.1033	0.1035	0.1102	0.1048	0.1131	0.1138	0.1156	0.1150	0.1182	**0.1216**	$7.7e^{-6}$
	NDCG@20	0.0856	0.0654	0.0650	0.0816	0.0852	0.0901	0.0873	0.0901	0.0943	0.0881	0.0971	0.0959	0.0990	0.0977	0.0981	**0.1015**	$4.9e^{-3}$
	Recall@40	0.1602	0.1203	0.1325	0.1553	0.1650	0.1766	0.1653	0.1662	0.1757	0.1723	0.1819	0.1776	0.1847	0.1799	0.1797	**0.1867**	$1.3e^{-2}$
	NDCG@40	0.1016	0.0754	0.0794	0.0980	0.1031	0.1085	0.1032	0.1062	0.1113	0.1068	0.1150	0.1122	0.1176	0.1156	0.1139	**0.1182**	$2.4e^{-1}$

Table 3: Ablation study on key components of AdaGCL.

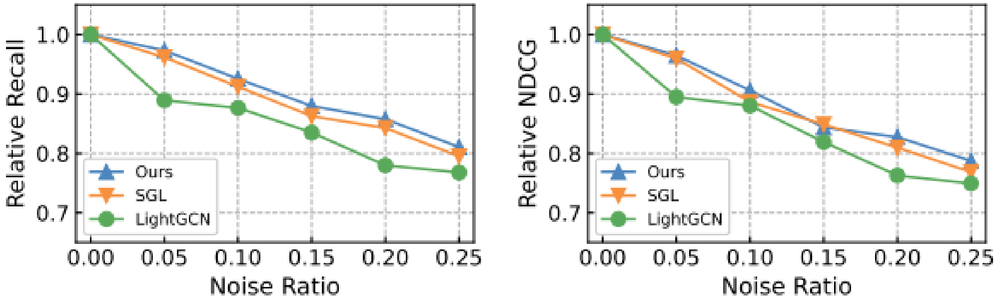
Category	Data Variants	Last.FM Recall	Last.FM NDCG	Yelp Recall	Yelp NDCG	BeerAdvocate Recall	BeerAdvocate NDCG
Adaptive	w/o Task	0.2562	0.1868	0.0849	0.0425	0.1212	0.1010
	Gen+Gen	0.2494	0.1819	0.0853	0.0429	0.1187	0.0992
Random	EdgeD	0.2476	0.1794	0.0852	0.0424	0.1163	0.0964
AdaGCL		0.2603	0.1911	0.0873	0.0439	0.1216	0.1015

using the random edge drop augmentation

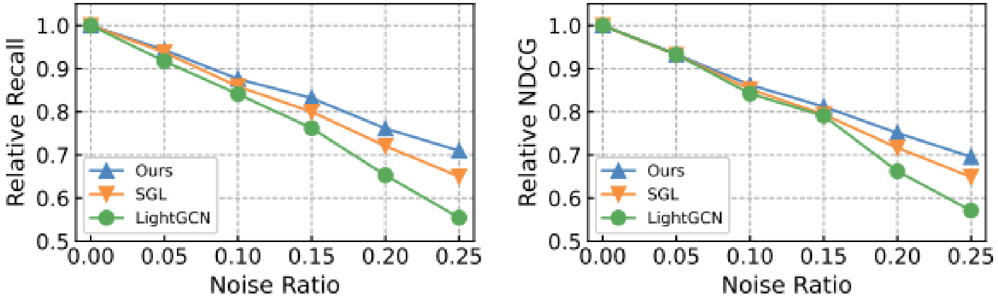
replace the denoising view generator with an identical VGAE-based generator

replace the task-aware optimization with the original reconstruction objective

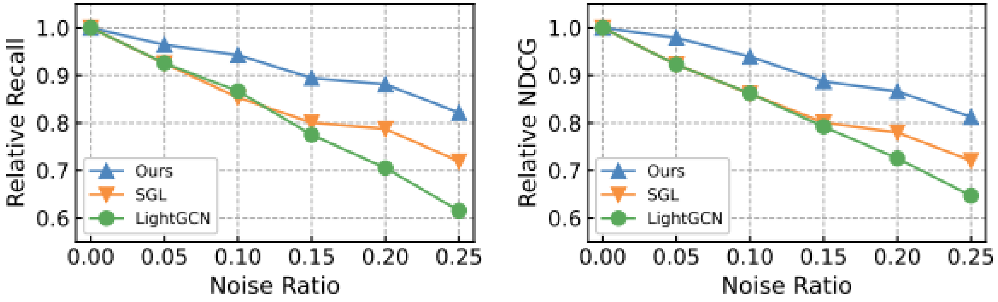
Experiment



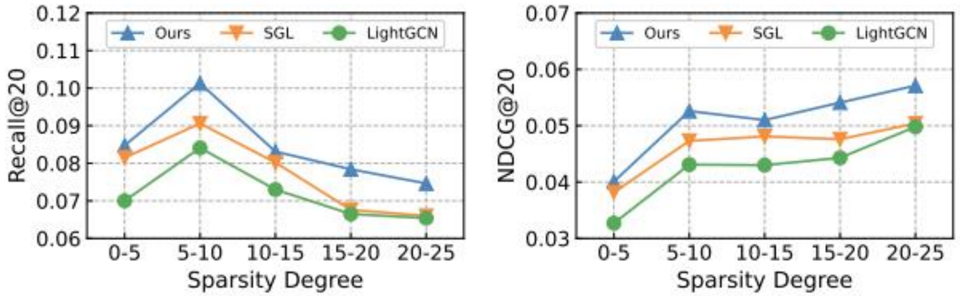
(a) Lasat.FM data



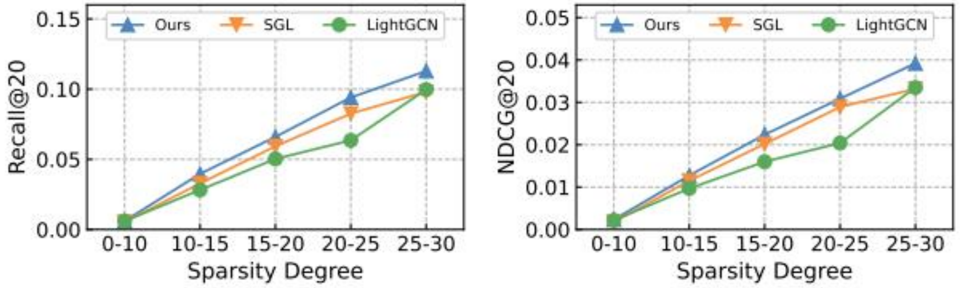
(b) Yelp data



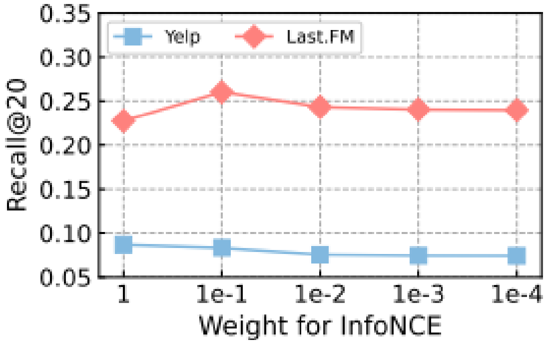
(c) BeerAdvocate data



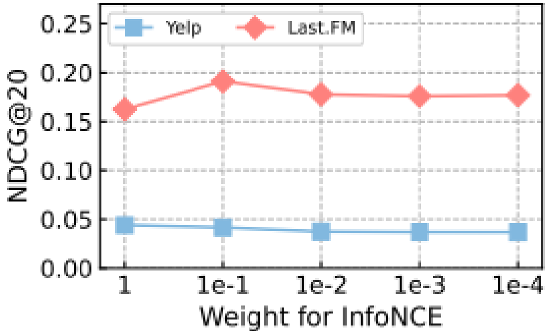
(a) Performance *w.r.t.* user interaction numbers



(b) Performance *w.r.t.* item interaction numbers

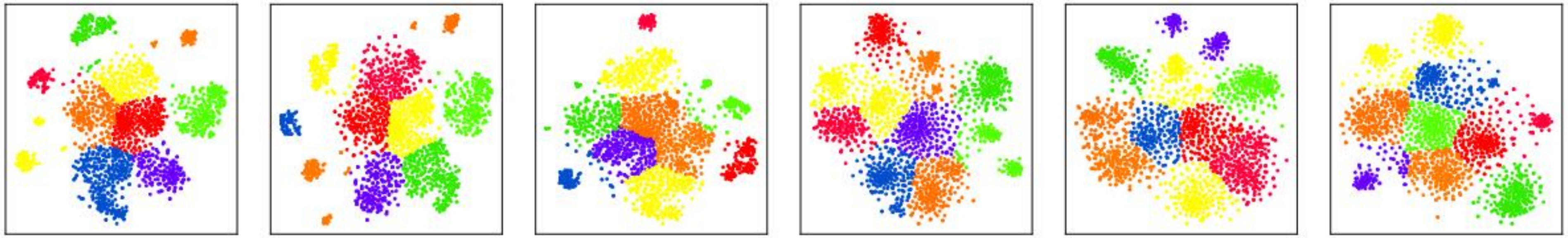


(a) Recall@20



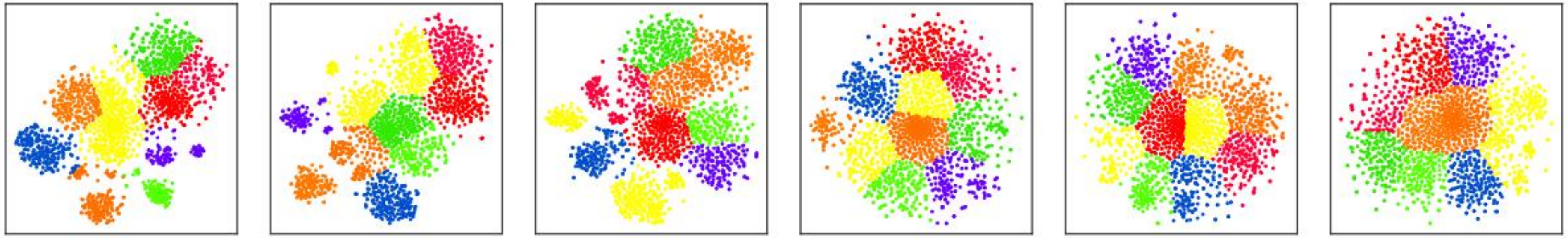
(b) NDCG@20

Experiment



(a) Main View (b) CL View 1 (c) CL View 2 (d) Noisy Main View (e) Noisy CL View 1 (f) Noisy CL View 2

Figure 6: View embedding visualization for AdaGCL.



(a) Main View (b) CL View 1 (c) CL View 2 (d) Noisy Main View (e) Noisy CL View 1 (f) Noisy CL View 2

Figure 7: View embedding visualization for SGL.

View 1 and View 2 are generated by the graph generative model and the graph denoising model, respectively.

Thank you!