Generative-Contrastive Graph Learning for Recommendation

Yonghui Yang
Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology
yyh.hfut@gmail.com

Zhengwei Wu
Ant Group
zejun.wzw@antfin.com

Le Wu*
Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology
lewu.ustc@gmail.com

Kun Zhang
Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology
zhang1028kun@gmail.com

Richang Hong
Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology
hongrc.hfut@gmail.com

Zhiqiang Zhang
Ant Group
lingyao.zzq@antfin.com

Jun Zhou
Ant Group
jun.zhoujun@antfin.com

Meng Wang
Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology
Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
eric.mengwang@gmail.com

（SIGIR-2023) github.com/yimutianyang/SIGIR23-VGCL

Reported by Ke Gan

1. Introduction
2. Approach
3. Experiments

Introduction

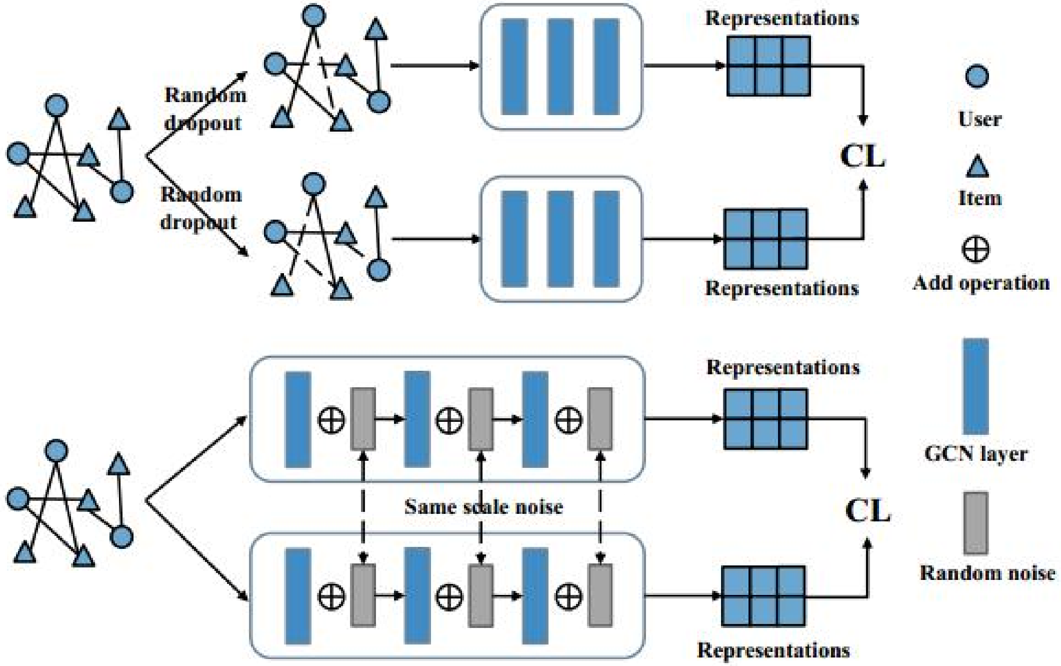


Figure 1: Graph contrastive learning paradigms with structure and feature data augmentation.

structure augmentation randomly perturb graph structure to obtain two augmented views

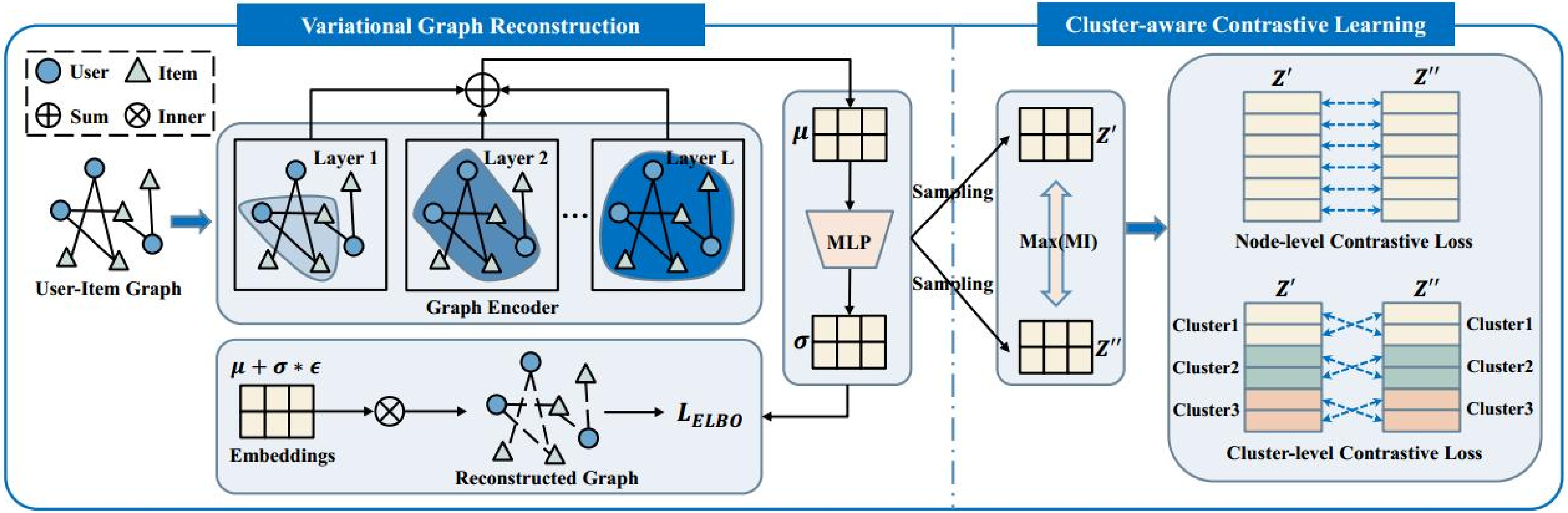
$$\mathbf{E}' = \mathcal{E}(\mathcal{G}', \mathbf{E}^0), \mathbf{E}'' = \mathcal{E}(\mathcal{G}'', \mathbf{E}^0), \tag{7}$$

feature augmentation adds random noises into node embeddings,then generate contrastive representations with GNNs

$$\mathbf{E}' = \mathcal{E}(\mathbf{E}^0, \epsilon\delta'), \mathbf{E}'' = \mathcal{E}(\mathbf{E}^0, \epsilon\delta''), \tag{8}$$

$\delta', \delta'' \sim U(0,1)$ are uniform noises, ϵ is the amplitude that controls noise scale.

Approach



$$A = \begin{bmatrix} 0^{M\times M} & R \\ R^T & 0^{N\times N} \end{bmatrix}. \tag{1}$$

$$E^l = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} E^{l-1}, \tag{2}$$

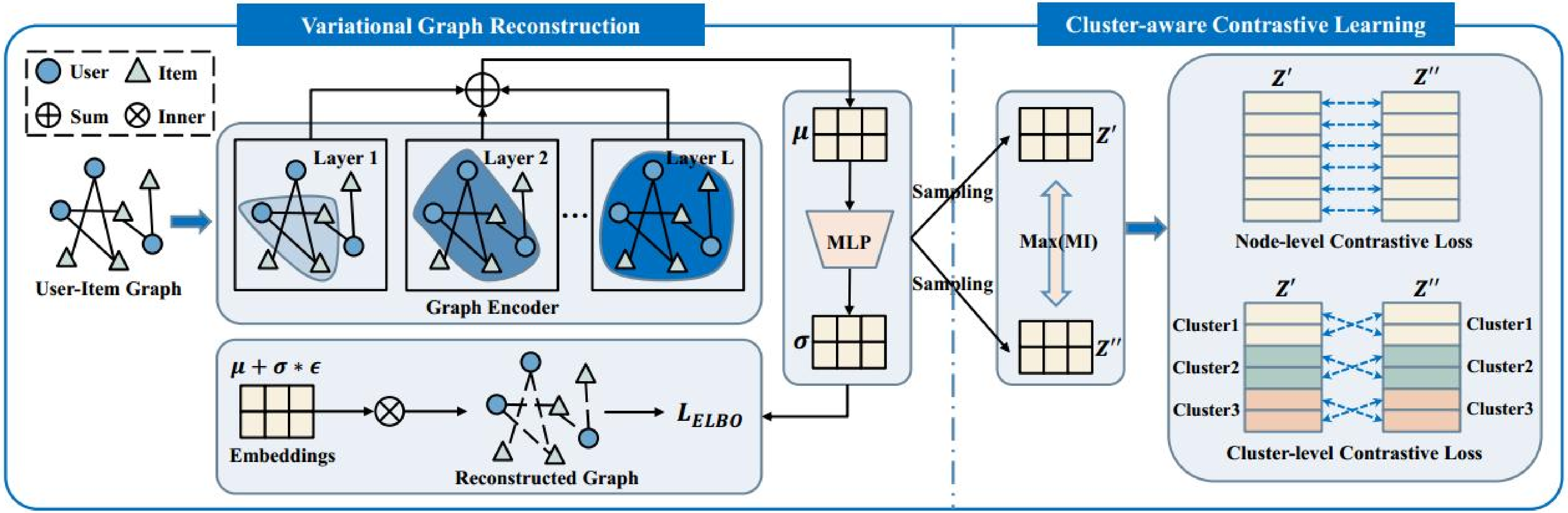
$$E = Readout(E^0, E^1, ..., E^L). \tag{3}$$

$$\mathcal{L}_{rec} = \sum_{a=0}^{M-1} \sum_{(i,j)\in D_a} -log\sigma(\hat{r}_{ai} - \hat{r}_{aj}) + \lambda ||E^0||^2, \tag{4}$$

$$\mathcal{L}_{cl} = \sum_{i\in\mathcal{B}} -log \frac{exp(e_i'^T e_i''/\tau)}{\sum_{j\in\mathcal{B}} exp(e_i'^T e_j''/\tau)}, \tag{6}$$

$$\mathcal{L} = \mathcal{L}_{rec} + \alpha \mathcal{L}_{cl}, \tag{5}$$

Approach



Graph Inference.

Gaussian distribution $q_\phi(\mathbf{z}_i|\mathbf{A},\mathbf{E}^0) = \mathcal{N}(\mathbf{z}_i|\mu_\phi(i), diag(\sigma^2_\phi(i)))$

$$\mu = GNN(\mathbf{A},\mathbf{E}^0,\phi_\mu), \sigma = GNN(\mathbf{A},\mathbf{E}^0,\phi_\sigma), \tag{11}$$

$$\mu_i^l = \sum_{j\in\mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_i|}\sqrt{|\mathcal{N}_j|}} \mu_i^{l-1}, \tag{12}$$

$$\mu = \frac{1}{L}\sum_{l=1}^{L} \mu^l, \sigma = MLP(\mu), \tag{13}$$

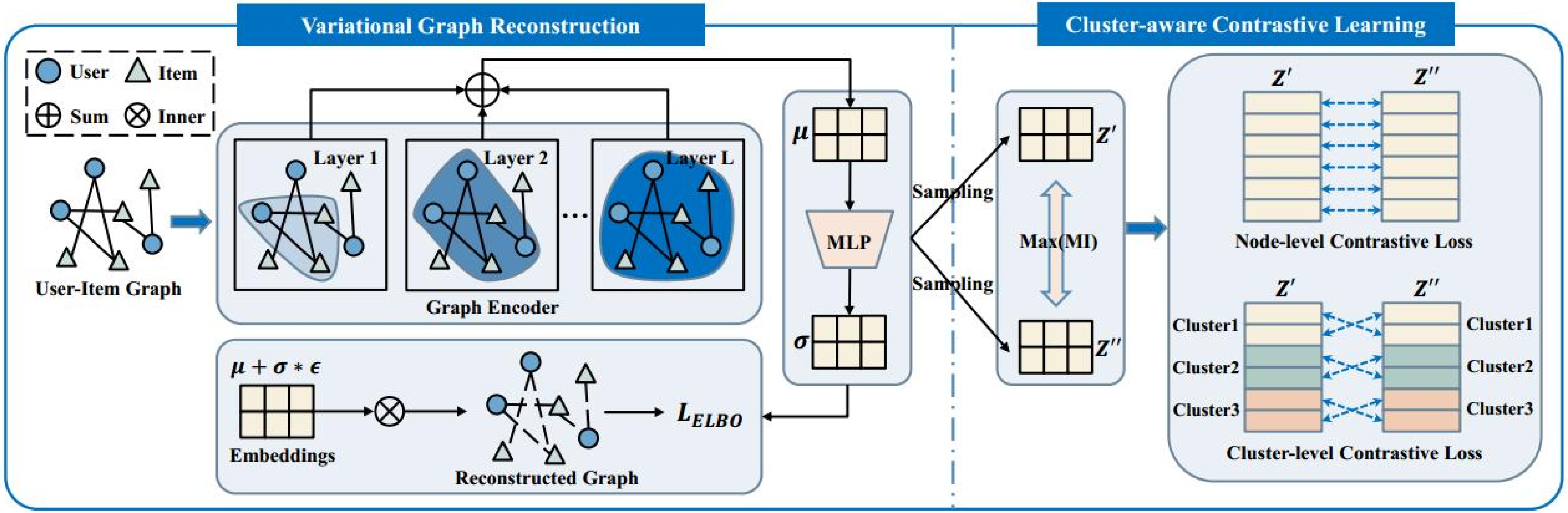
$$\mathbf{z}_i = \mu_i + \sigma_i \cdot \varepsilon, \tag{14}$$

Graph Generation.

$$p(\mathbf{A}|\mathbf{Z}) = \prod_{i=0}^{M+N-1}\prod_{j=0}^{M+N-1} p(\mathbf{A}_{ij}|\mathbf{z}_i,\mathbf{z}_j). \tag{15}$$

$$p(\mathbf{A}_{ij} = 1|\mathbf{z}_i,\mathbf{z}_j) = \sigma(\mathbf{z}_i^T\mathbf{z}_j), \tag{16}$$

Approach



Cluster-aware Contrastive Learning

$$\mathbf{z}'_i = \mu_i + \sigma_i \cdot \varepsilon', \tag{17}$$

$$\mathbf{z}''_i = \mu_i + \sigma_i \cdot \varepsilon'', \tag{18}$$

$$\mathcal{L}_N^U = \sum_{a\in\mathcal{B}_u} -log \frac{exp(\mathbf{z}_a'^{T}\mathbf{z}_a''/\tau_1)}{\sum_{b\in\mathcal{B}_u} exp(\mathbf{z}_a'^{T}\mathbf{z}_b''/\tau_1)}, \tag{19}$$

$$\mathcal{L}_N^I = \sum_{i\in\mathcal{B}_i} -log \frac{exp(\mathbf{z}_i'^{T}\mathbf{z}_i''/\tau_1)}{\sum_{j\in\mathcal{B}_i} exp(\mathbf{z}_i'^{T}\mathbf{z}_j''/\tau_1)}, \tag{20}$$

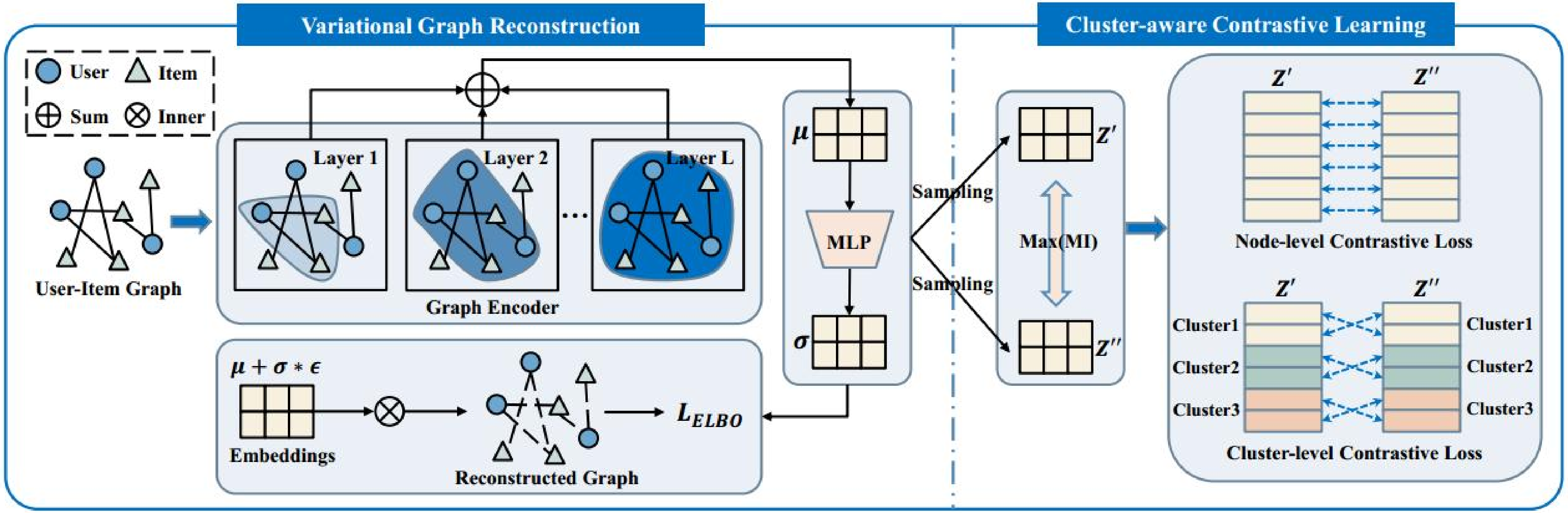
$$p(a,b) = \sum_{k=0}^{K_u-1} p(\mathbf{c}_k^u|\mathbf{z}_a)p(\mathbf{c}_k^u|\mathbf{z}_b), \tag{21}$$

$$p(i,j) = \sum_{h=0}^{K_i-1} p(\mathbf{c}_h^i|\mathbf{z}_i)p(\mathbf{c}_h^i|\mathbf{z}_j), \tag{22}$$

$$\mathcal{L}_C^U = \sum_{a\in\mathcal{B}_u} \frac{-1}{SP(a)} log(\frac{\sum_{b\in\mathcal{B}_u, b!=a} p(a,b) exp(\mathbf{z'}_a^T\mathbf{z''}_b/\tau_2)}{\sum_{b\in\mathcal{B}_u, b!=a} exp(\mathbf{z'}_a^T\mathbf{z''}_b/\tau_2)}), \tag{23}$$

$$\mathcal{L}_C^I = \sum_{i\in\mathcal{B}_i} \frac{-1}{SP(i)} log(\frac{\sum_{j\in\mathcal{B}_i, j!=i} p(i,j) exp(\mathbf{z'}_i^T\mathbf{z''}_j/\tau_2)}{\sum_{j\in\mathcal{B}_i, j!=i} exp(\mathbf{z'}_i^T\mathbf{z''}_j/\tau_2)}), \tag{24}$$

Approach



$$\mathcal{L}_{cl} = \mathcal{L}_N + \gamma \mathcal{L}_C, \tag{25}$$

$$\mathcal{L}_{ELBO} = -\mathbb{E}_{Z\sim q_\phi(Z|A,E^0)}[log(p_\theta(A|Z))] + KL[q_\phi(Z|A,E^0)||p(Z)]. \tag{26}$$

$$\mathbb{E}_{Z\sim q_\phi(Z|A,E^0)}[log(p_\theta(A|Z))] = \sum_{a=0}^{M-1} \sum_{(i,j)\in D_a} -log\sigma(\hat{r}_{ai} - \hat{r}_{aj})), \tag{27}$$

$$min\mathcal{L} = \mathcal{L}_{ELBO} + \alpha\mathcal{L}_{cl} + \lambda||E^0||^2, \tag{28}$$

Experiment

Table 1: The statistics of three datasets.

Datasets	Users	Items	Interactions	Density
Douban-Book	13,024	22,347	792,062	0.272%
Dianping	59,426	10,224	934,334	0.154%
Movielens-25M	92,901	8,826	2,605,952	0.318%

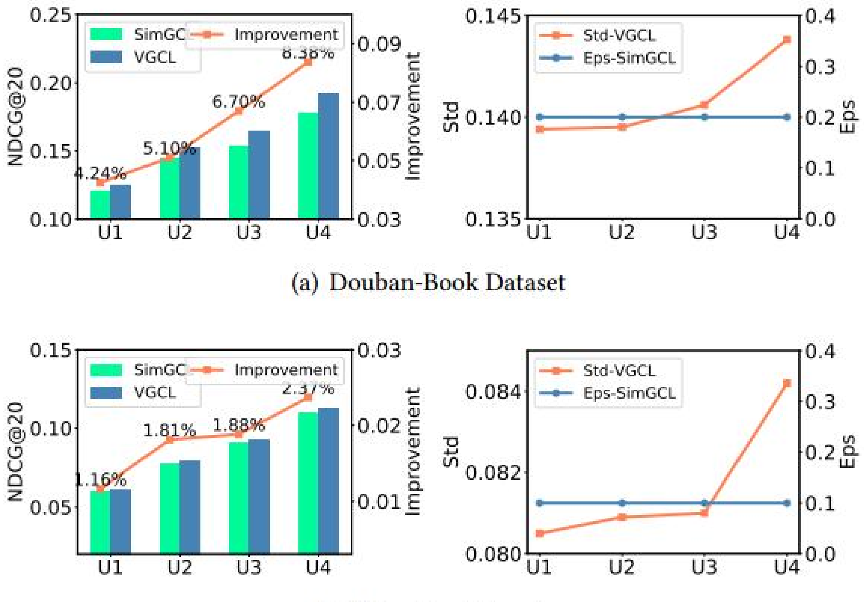
Table 2: Recommendation performances on three datasets. The best-performing model on each dataset and metrics are highlighted in bold, and the second-best model is underlined.

Models	Douban-Book				Dianping				Movielens-25M			
	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20
BPR-MF	0.0869	0.0949	0.1296	0.1045	0.0572	0.0443	0.0934	0.0557	0.2152	0.2011	0.3163	0.2343
LightGCN	0.1042	0.1195	0.1516	0.1278	0.0679	0.0536	0.1076	0.0660	0.2258	0.2192	0.3263	0.2509
Multi-VAE	0.0941	0.1073	0.1376	0.1155	0.0645	0.0508	0.1046	0.0632	0.2188	0.2101	0.3185	0.2418
CVGA	0.1058	0.1305	0.1492	0.1359	0.0719	0.0562	0.1128	0.0690	0.2390	0.2306	0.3454	0.2641
SGL-ED	0.1103	0.1357	0.1551	0.1419	0.0719	0.0560	0.1111	0.0686	0.2298	0.2239	0.3274	0.2541
NCL	0.1121	0.1377	0.1576	0.1439	0.0727	0.0571	0.1124	0.0701	0.2281	0.2222	0.3274	0.2531
SimGCL	<u>0.1218</u>	<u>0.1470</u>	<u>0.1731</u>	<u>0.1540</u>	<u>0.0768</u>	<u>0.0606</u>	<u>0.1208</u>	<u>0.0743</u>	<u>0.2428</u>	<u>0.2356</u>	<u>0.3491</u>	<u>0.2690</u>
VGCL	**0.1283**	**0.1564**	**0.1829**	**0.1638**	**0.0778**	**0.0616**	**0.1234**	**0.0757**	**0.2463**	**0.2400**	**0.3507**	**0.2725**

Experiment

Table 3: Ablation study of VGCL, VGCL-w/o C denotes without cluster-level contrastive loss and VGCL-w/o V denotes without the variational graph reconstruction part.

Models	Douban-Book		Dianping		Movielens-25M	
	R@20	N@20	R@20	N@20	R@20	N@20
LightGCN	0.1512(-)	0.1271(-)	0.1076(-)	0.0660(-)	0.3263(-)	0.2509(-)
SimGCL	0.1731(14.48%)	0.1540(+21.16%)	0.1208(+12.27%)	0.0743(+12.58%)	0.3491(+6.99%)	0.2690(+7.21%)
VGCL-w/o C	0.1776(+17.46%)	0.1575(+23.92%)	0.1222(+13.57%)	0.0750(+13.64%)	0.3477(+6.56%)	0.2705(+7.81%)
VGCL-w/o V	0.1722(+13.89%)	0.1547(+21.72%)	0.1218(+13.20%)	0.0746(+13.03%)	0.3493(+7.05%)	0.2702(+7.69%)
VGCL	**0.1829(+20.97%)**	**0.1638(+28.87%)**	**0.1233(+14.59%)**	**0.0756(+14.55%)**	**0.3507(+7.48%)**	**0.2725(+8.61%)**



(a) Douban-Book Dataset

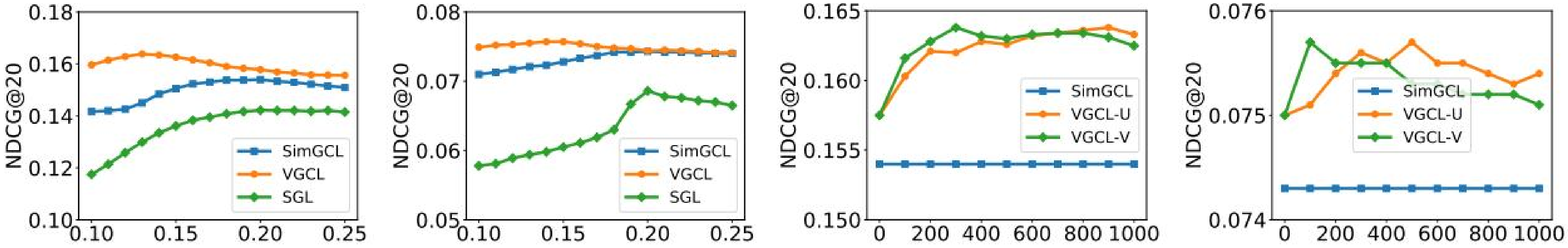
(b) Dianping Dataset

Figure 3: Performance comparisons under different user groups.

Table 4: Performance on different graph inference layer L.

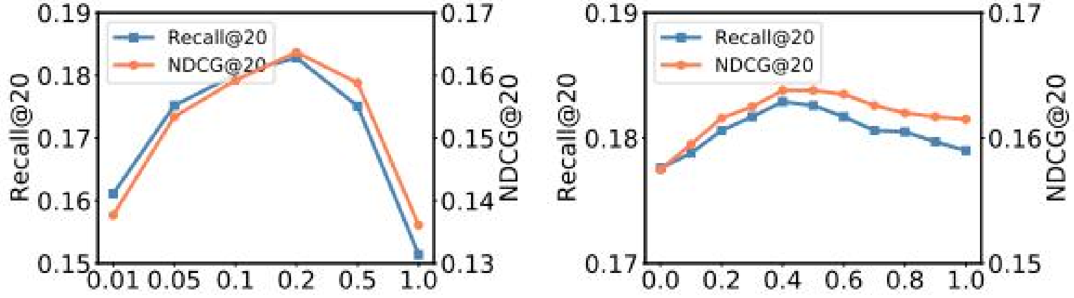
Layers	Douban-Book		Dianping	
	Recall@20	NDCG@20	Recall@20	NDCG@20
L=1	0.1750	0.1555	0.1197	0.0733
L=2	**0.1829**	**0.1638**	0.1229	0.0751
L=3	0.1808	0.1618	**0.1234**	**0.0757**
L=4	0.1793	0.1605	0.1233	0.0752

Experiment



(a) τ on Douban-Book (b) τ on Dianping (c) K_u, K_i on Douban-Book (d) K_u, K_i on Dianping

Figure 4: Performance comparisons w.r.t different temperature τ, prototype number K_u and K_i.



(a) α on Douban-Book (b) γ on Douban-Book

Figure 5: Performance comparisons under different contrastie loss weights α and γ.

Thank you!