Contrastive Graph Structure Learning via Information Bottleneck for Recommendation

Chunyu Wei[1]*, Jian Liang[1]*, Di Liu[1], Fei Wang[2]
[1]Alibaba Group, China
[2]Department of Population Health Sciences, Weill Cornell Medicine, USA
weicy15@icloud.com
{xuelang.lj, wendi.ld}@alibaba-inc.com
few2001@med.cornell.edu

NIPS 2022
Code: https://github.com/weicy15/CGI

Reported by liang li

Motivation

Details:

- The graph structures generated from before may be suboptimal.
- Maximizing the mutual information in the contrastive learning may push the representations of different views to capture wrong information.

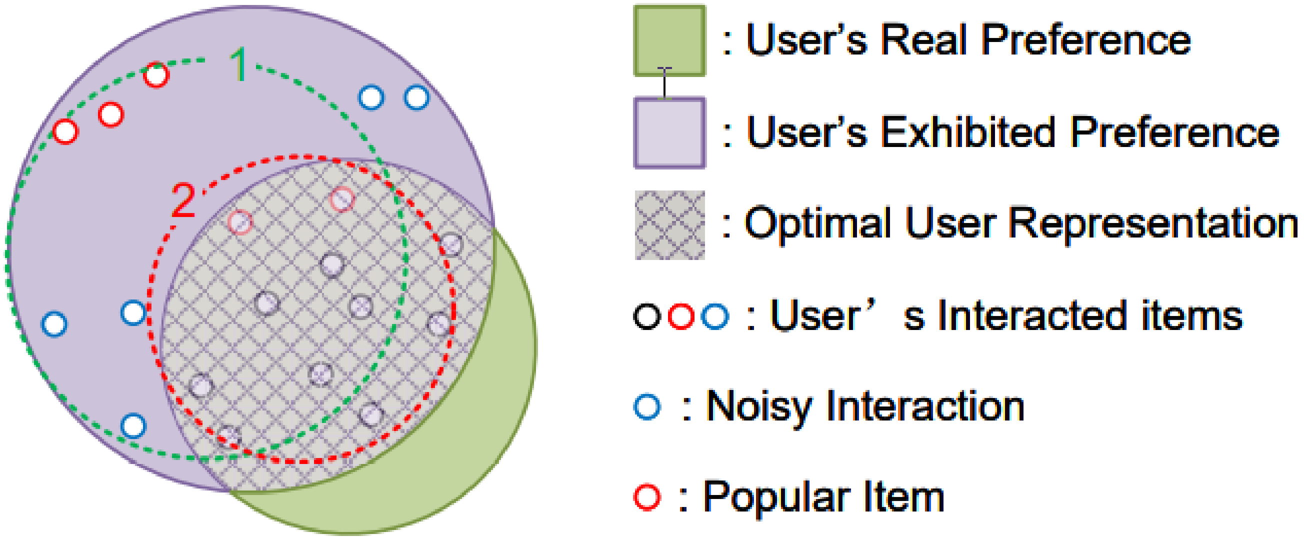


Figure 1: A possible illustration of some user's interactions and preference. Dotted circles denote possible augmentation representations.

Problem Statement

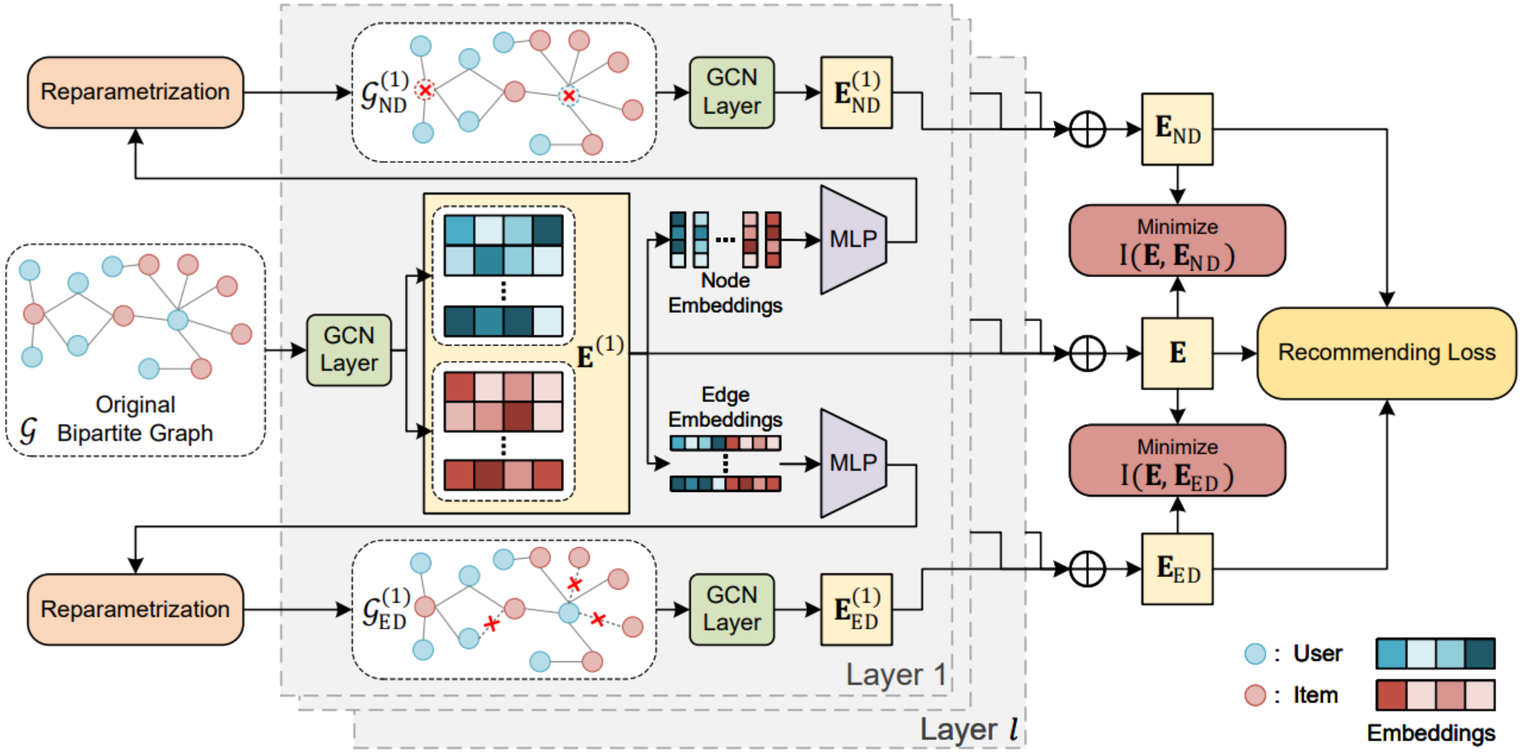


Figure 2: The overview the CGI framework. We integrate both the node-dropping and edge-dropping views together for a more comprehensive representation, though they can be applied separately.

$$\mathcal{U} = \{u_1, u_2, \ldots, u_m\}$$

$$\mathcal{I} = \{i_1, i_2, \ldots, i_n\}$$

$$\mathbf{R} \in \mathbb{R}^{m \times n}$$

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\} \quad \mathcal{V} = \mathcal{U} \cup \mathcal{I}$$

$$\mathcal{E} = \{e_{ui} | r_{ui} = 1, u \in \mathcal{U}, i \in \mathcal{I}\}$$

$$\mathbf{D}_{\mathcal{G}} \in \mathbb{N}^{(m+n)\times(m+n)}$$

Method

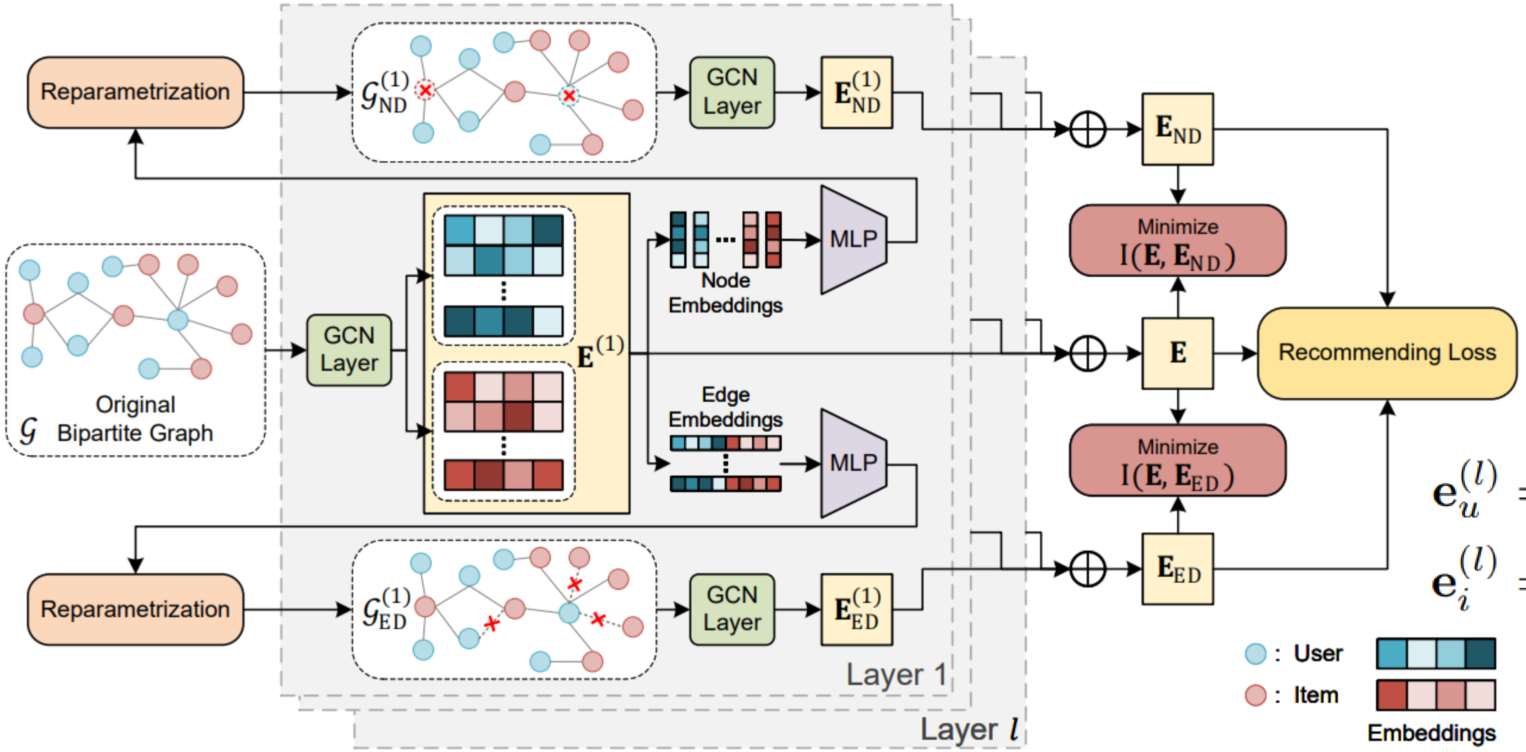


Figure 2: The overview the CGI framework. We integrate both the node-dropping and edge-dropping views together for a more comprehensive representation, though they can be applied separately.

$$\max_{\mathbf{Z}} I(\mathbf{Y};\mathbf{Z}) - \beta I(\mathbf{X};\mathbf{Z}), \quad (1)$$

$$\mathbf{A}_{\mathcal{G}} = \begin{bmatrix} 0 & \mathbf{R} \\ \mathbf{R}^T & 0 \end{bmatrix}. \quad (2)$$

$$\mathbf{E}^{(l)} = GCN(\mathbf{E}^{(l-1)}, \mathcal{G}), \quad (3)$$

$$\mathbf{e}_u^{(l)} = f_{combine}^{(l)}(\mathbf{e}_u^{(l-1)}, f_{aggregate}^{(l)}(\{\mathbf{e}_i^{(l)} | i \in \mathcal{N}_u\})), \quad (4)$$

$$\mathbf{e}_i^{(l)} = f_{combine}^{(l)}(\mathbf{e}_i^{(l-1)}, f_{aggregate}^{(l)}(\{\mathbf{e}_u^{(l)} | u \in \mathcal{N}_i\})), \quad (5)$$

Method

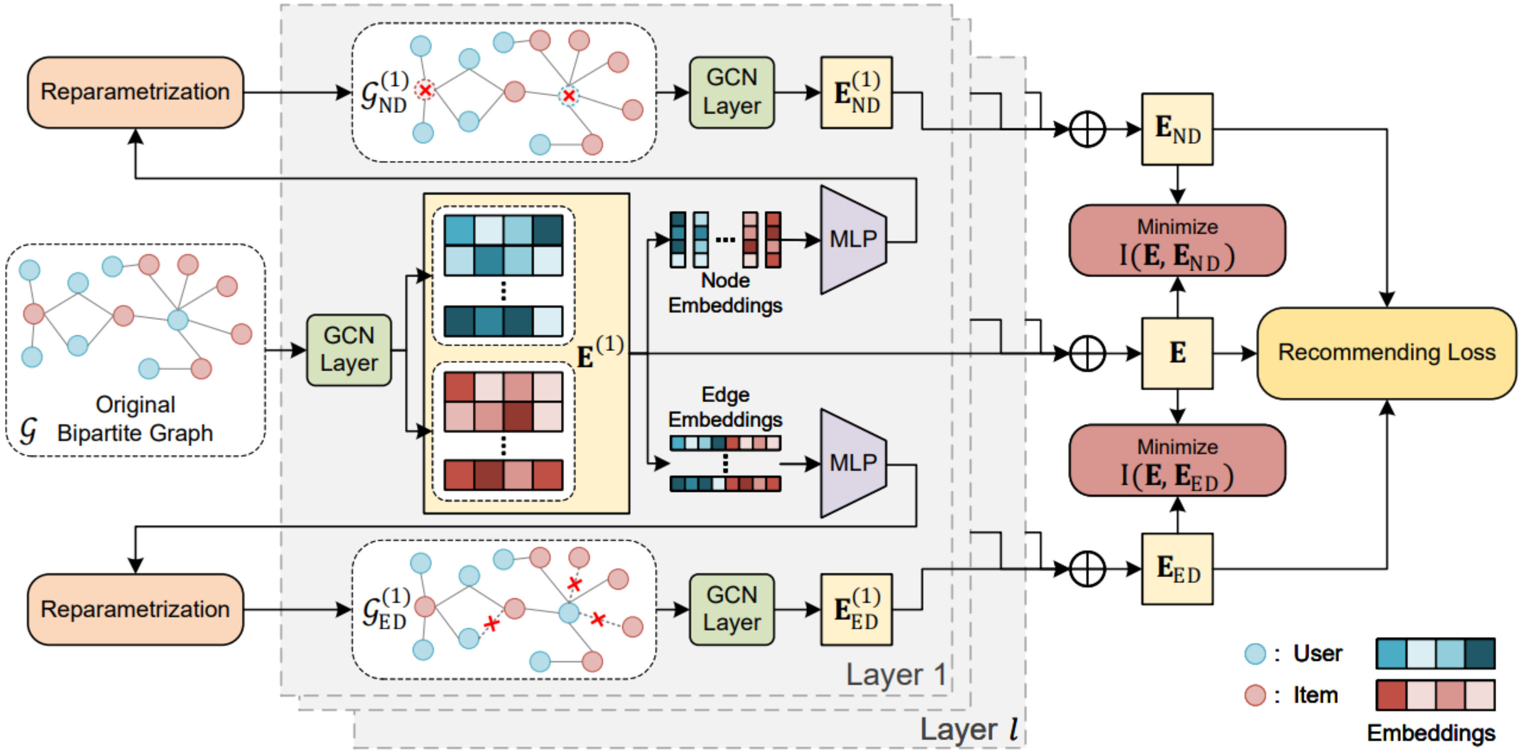


Figure 2: The overview the CGI framework. We integrate both the node-dropping and edge-dropping views together for a more comprehensive representation, though they can be applied separately.

$$\mathbf{e} = f_{readout}(\{\mathbf{e}^{(l)} | l = 0, 1, \dots, L\}). \quad (6)$$

$$\mathbf{E}^{(l)} = (\mathbf{D}_{\mathcal{G}}^{-\frac{1}{2}} \mathbf{A}_{\mathcal{G}} \mathbf{D}_{\mathcal{G}}^{-\frac{1}{2}})\mathbf{E}^{(l-1)}, l \in \mathbb{N}^{+}, \quad (7)$$

Method

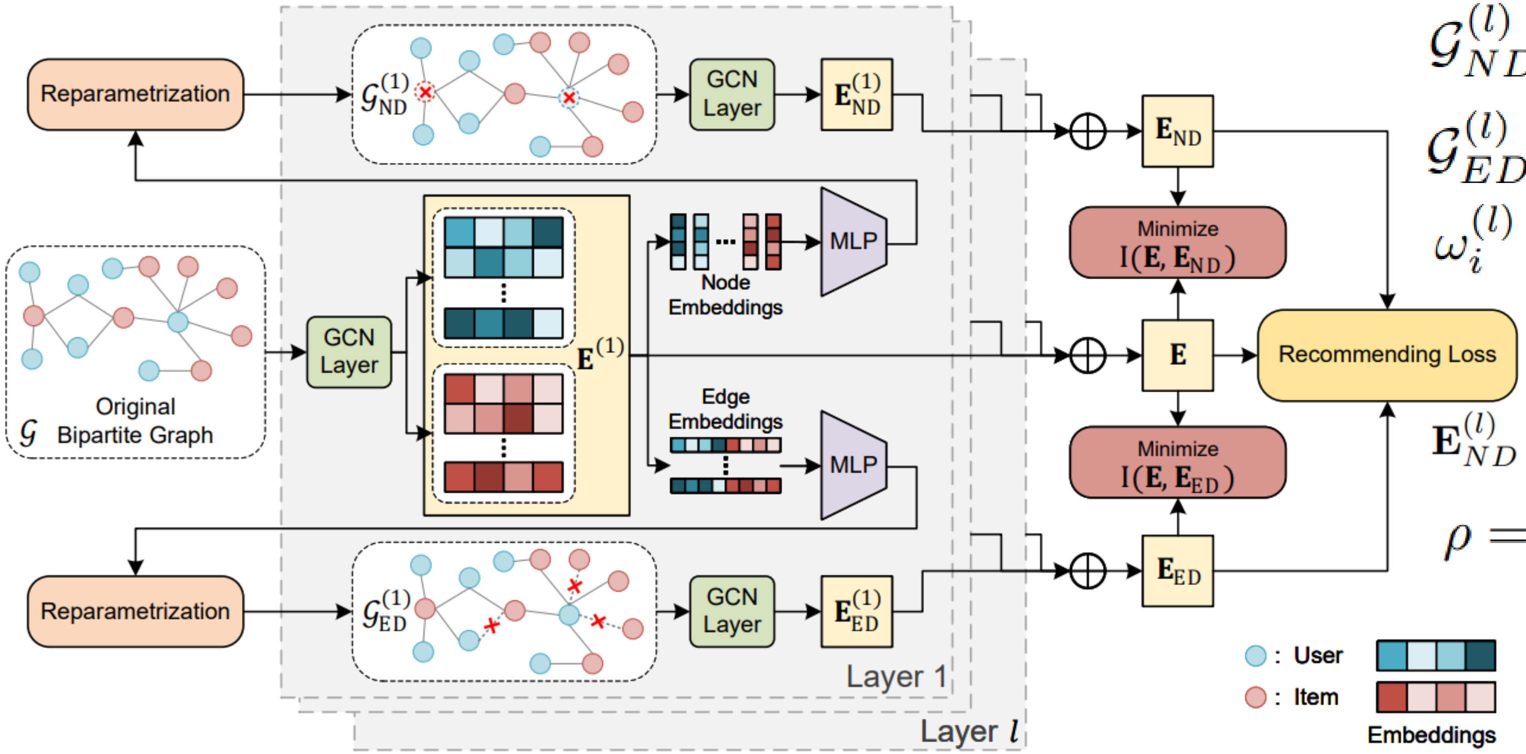


Figure 2: The overview the CGI framework. We integrate both the node-dropping and edge-dropping views together for a more comprehensive representation, though they can be applied separately.

$$\mathcal{G}_{ND}^{(l)} = \{\{v_i \odot \rho_i^{(l)} \mid v_i \in \mathcal{V}\}, \mathcal{E}\}, (8)$$

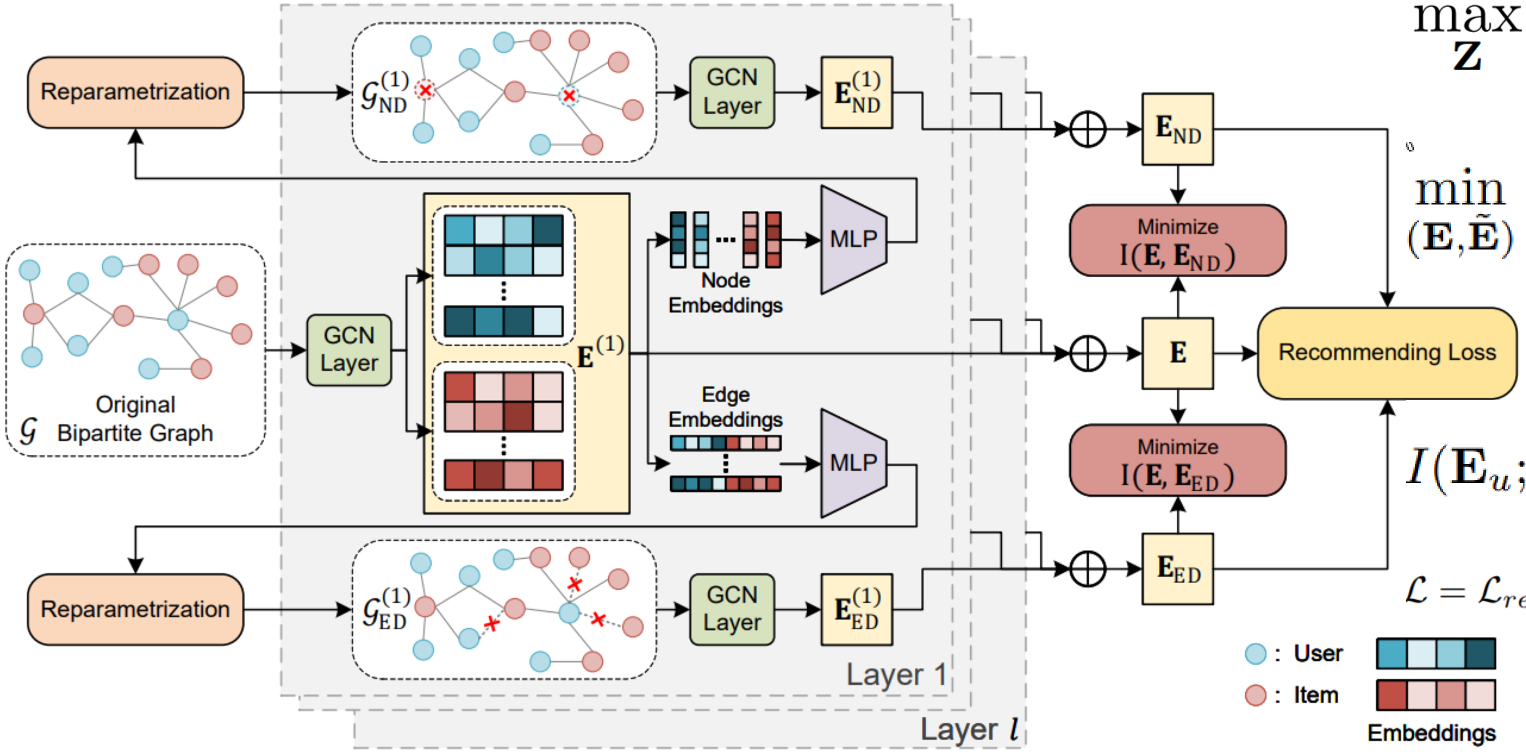
$$\mathcal{G}_{ED}^{(l)} = \{\mathcal{V}, \{e_{ij} \odot \rho_{ij}^{(l)} \mid e_{ij} \in \mathcal{E}\}\}, (9)$$

$$\omega_i^{(l)} = MLP(\mathbf{e}_i^{(l)}); \ \ \omega_{ij}^{(l)} = MLP([\mathbf{e}_i^{(l)}; \mathbf{e}_j^{(l)}]). (10)$$

$$\mathbf{E}_{ND}^{(l)} = GCN(\mathbf{E}_{ND}^{(l-1)}, \mathcal{G}_{ND}^{(l)}), \ \ \mathbf{E}_{ED}^{(l)} = GCN(\mathbf{E}_{ED}^{(l-1)}, \mathcal{G}_{ED}^{(l)}), (12)$$

$$\rho = \sigma((\log \epsilon - \log(1-\epsilon) + \omega)/\tau), (11)$$

Method



$$\max_{\mathbf{Z}} I(\mathbf{Y};\mathbf{Z}) - \beta I(\mathbf{X};\mathbf{Z}), \quad (1)$$

$$\min_{(\mathbf{E},\tilde{\mathbf{E}})} \tilde{\mathcal{L}}_{rec} + I(\mathbf{E};\tilde{\mathbf{E}}), \quad (13)$$

$$I(\mathbf{E}_u;\tilde{\mathbf{E}}_u) = \sum_{v_i\in\mathcal{U}} \log \frac{exp(s(\mathbf{e}_i,\tilde{\mathbf{e}}_i)/\tau')}{\sum_{v_j\in\mathcal{U}} exp(s(\mathbf{e}_i,\tilde{\mathbf{e}}_j)/\tau')}, \quad (14)$$

$$\mathcal{L} = \mathcal{L}_{rec} + \mathcal{L}^{ND}_{rec} + \mathcal{L}^{ED}_{rec} + \lambda(I(\mathbf{E},\mathbf{E}_{ND}) + I(\mathbf{E},\mathbf{E}_{ED})) + \beta\|\Theta\|_2^2, \quad (15)$$

$$\mathcal{L}_{rec} = \sum_{(u,i,j)\in\mathcal{O}} -ln\sigma(\hat{r}_{ui} - \hat{r}_{uj}),$$

Figure 2: The overview the CGI framework. We integrate both the node-dropping and edge-dropping views together for a more comprehensive representation, though they can be applied separately.

Experiments

Table 1: Comparison among models. Boldface denotes the highest score and underline indicates the best result of the baselines.

Model	Yelp2018				MovieLens-1M			
	NDCG@10	RECALL@10	NDCG@20	RECALL@20	NDCG@10	RECALL@10	NDCG@20	RECALL@20
BPRMF	0.0138	0.0209	0.0191	0.0373	0.1225	0.1376	0.1407	0.1882
NCF	0.0224	0.0356	0.0289	0.0566	0.1430	0.1546	0.1576	0.2027
NGCF	0.0242	0.0384	0.0319	0.0629	0.1462	0.1651	0.1667	0.2285
LightGCN	0.0344	0.0530	0.0445	0.0850	0.1696	0.1865	0.1863	0.2420
DNN+SSL	0.0217	0.0344	0.0286	0.0564	0.1096	0.1238	0.1250	0.1714
SGL	<u>0.0367</u>	<u>0.0552</u>	<u>0.0473</u>	<u>0.0891</u>	<u>0.1800</u>	<u>0.1965</u>	<u>0.1972</u>	<u>0.2520</u>
CGI	**0.0392**	**0.0584**	**0.0501**	**0.0932**	**0.1979**	**0.2180**	**0.2152**	**0.2772**
Improv.	+6.82%	+5.90%	+5.93%	+4.58%	+9.95%	+10.91%	+9.13%	+9.97%
p-value	1.29e-3	3.53e-3	7.00e-4	3.59e-4	8.89e-4	4.22e-4	4.83e-4	5.07e-5

Model	Douban			
	NDCG@10	RECALL@10	NDCG@20	RECALL@20
BPRMF	0.0496	0.0526	0.0516	0.0613
NCF	0.0694	0.0706	0.0659	0.0734
NGCF	0.0794	0.0823	0.0784	0.0897
LightGCN	0.0862	0.0876	0.0845	0.0940
DNN+SSL	0.0712	0.0738	0.0703	0.0804
SGL	<u>0.0912</u>	<u>0.0906</u>	<u>0.0910</u>	<u>0.1012</u>
CGI	**0.0991**	**0.1007**	**0.0979**	**0.1119**
Improv.	+8.69%	+11.18%	+7.55%	+10.55%
p-value	1.99e-3	4.40e-3	1.52e-4	1.60e-4

Experiments

Model	Yelp2018	
	NDCG@10	RECALL@10
LightGCN	0.0344	0.0530
CGI	**0.0392**	**0.0584**
SGL-ND	0.0356	0.0544
CGI-ND	0.0369	0.0569
SGL-ED	0.0367	0.0552
CGI-ED	0.0379	0.0579

Model	MovieLens-1M	
	NDCG@10	RECALL@10
LightGCN	0.1696	0.1865
CGI	**0.1979**	**0.2180**
SGL-ND	0.1765	0.1948
CGI-ND	0.1934	0.2119
SGL-ED	0.1800	0.1965
CGI-ED	0.1916	0.2088

Table 2: Comparison among models.

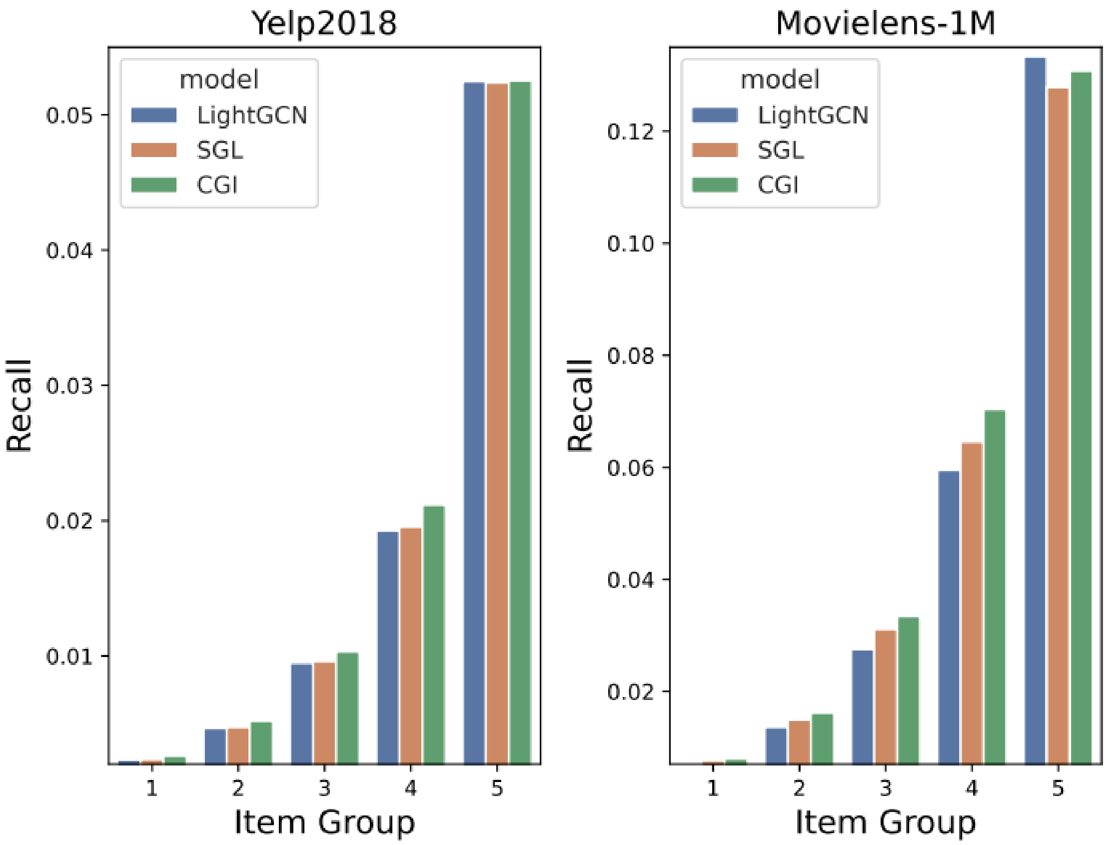


Figure 3: Performance of different item groups

Experiments

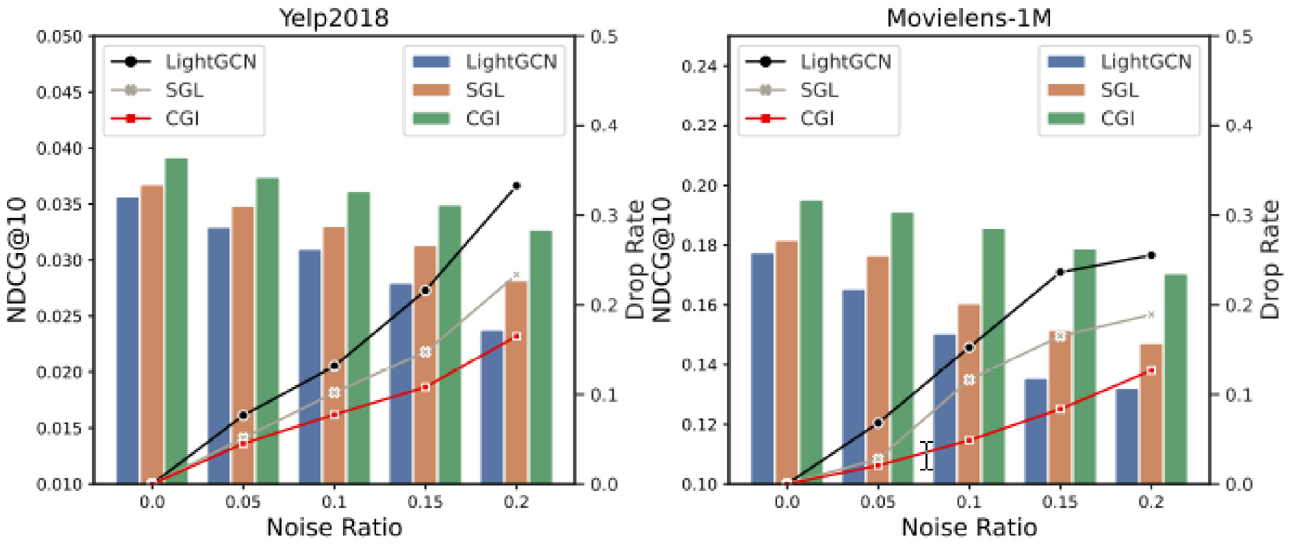


Figure 4: Performance comparison over different noise ratio. The bar represents the NDCG@10 and the line represent the performance degradation ratio.

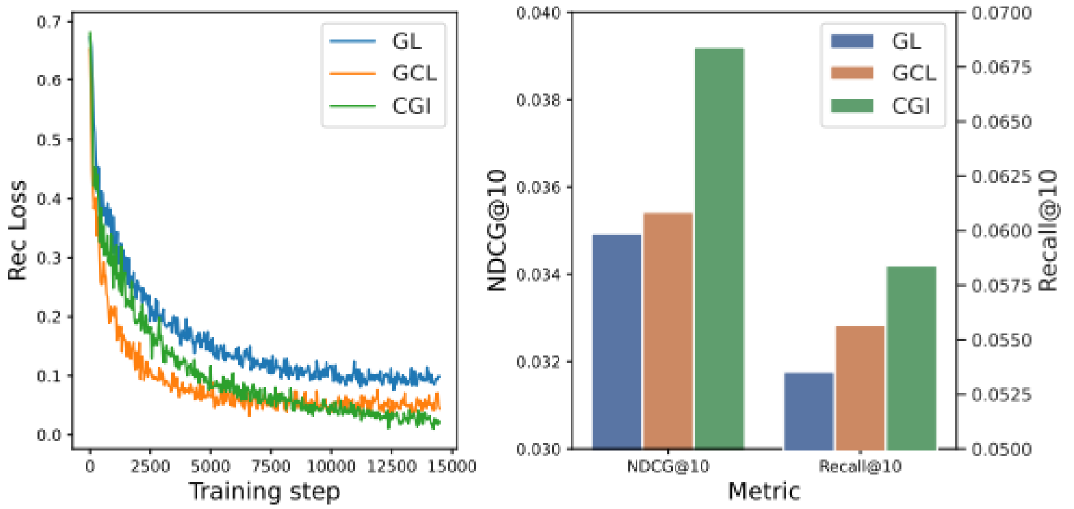


Figure 5: Effect of Information Bottleneck on Yelp2018

Experiments

Table 3: Performance with Other GNN variants.

Model	Yelp2018		Movielens-1M		Douban	
	NDCG@10	Recall@10	NDCG@10	Recall@10	NDCG@10	Recall@10
GC-MC	0.0214	0.0278	0.1350	0.1491	0.0671	0.0739
SGL+GC-MC	0.0218(+1.9%)	0.0281(+1.2%)	0.1412(+4.6%)	0.1577(+5.8%)	0.0687(+2.3%)	0.0762(+3.1%)
CGI+GC-MC	0.0218(+2.1%)	0.0282(+1.7%)	0.1422(+5.3%)	0.1585(+6.3%)	0.0687(+2.3%)	0.0765(+3.5%)
NGCF	0.0242	0.0384	0.1462	0.1651	0.0794	0.0823
SGL+NGCF	0.0260(+7.4%)	0.0418(+8.9%)	0.1609(+10.1%)	0.1871(+13.3%)	0.0833(+4.9%)	0.0857(+4.1%)
CGI+NGCF	0.0272(+12.5%)	0.0431(12.1%)	0.1660(%13.6%)	0.1937(+17.3%)	0.0840(+5.7%)	0.0875(+6.3%)

Thanks