Automated Self-Supervised Learning for Recommendation

Lianghao Xia
The University of Hong Kong
aka_xia@foxmail.com

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

Chunzhen Huang
Tencent
chunzhuang@tencent.com

Kangyi Lin
Tencent
plancklin@tencent.com

Tao Yu
The University of Hong Kong
tyu@cs.hku.hk

Ben Kao
The University of Hong Kong
kao@cs.hku.hk

https://github.com/HKUDS/AutoCF

（WWW-2022)

Leibniz-Institut
für Sozialwissenschaften

Reported by Ke Gan

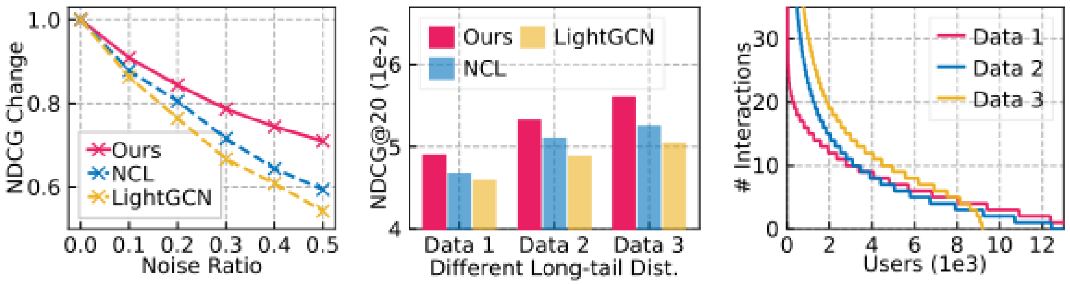
1. Introduction
2. Approach
3. Experiments

Introduction

rely on manually generated contrastive views for heuristic-based data augmentation

i) losing important structural information

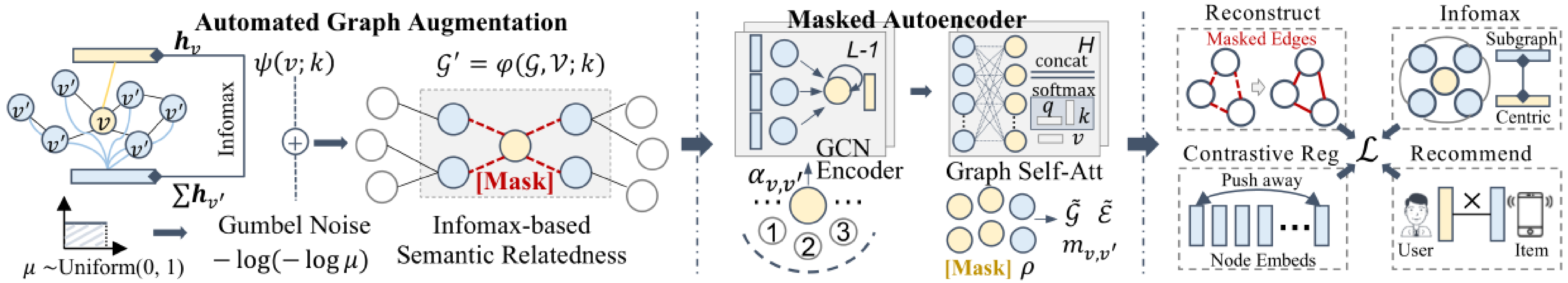
ii) keeping the noisy data



(a) Impact of data noise (b) Impact of long-tail dist (c) Dist. of datasets

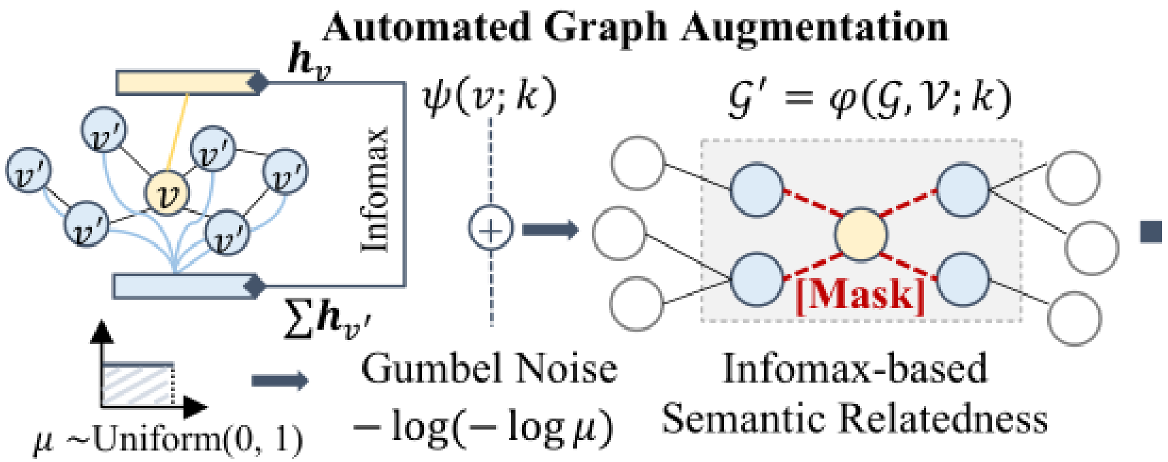
Figure 1: The influence of data noise and long-tail distributions on the prediction accuracy of different methods.

Approach



Automated Graph Augmentation
h_v
$\psi(v;k)$
$\mathcal{G}' = \varphi(\mathcal{G}, \mathcal{V}; k)$
Infomax
[Mask]
$\sum h_{v'}$
μ ~Uniform(0, 1)
Gumbel Noise
$-\log(-\log\mu)$
Infomax-based
Semantic Relatedness
Masked Autoencoder
L-1
GCN
Encoder
$\alpha_{v,v'}$
H
concat
softmax
q
k
v
Graph Self-Att
$\tilde{\mathcal{G}}$ $\tilde{\mathcal{E}}$
$m_{v,v'}$
[Mask] ρ
Reconstruct
Masked Edges
Infomax
Subgraph
Centric
Contrastive Reg
$\mathcal{L}$
Recommend
Push away
Node Embeds
User
Item

Approach



$$\varphi(\mathcal{G}, \mathcal{V}; k) = \left\{\mathcal{U}, \mathcal{I}, \mathcal{V}, \mathcal{E} \setminus \{(v_1, v_2) | v_1, v_2 \in \mathcal{N}_v^k, v \in \mathcal{V}\}\right\} \quad (3)$$

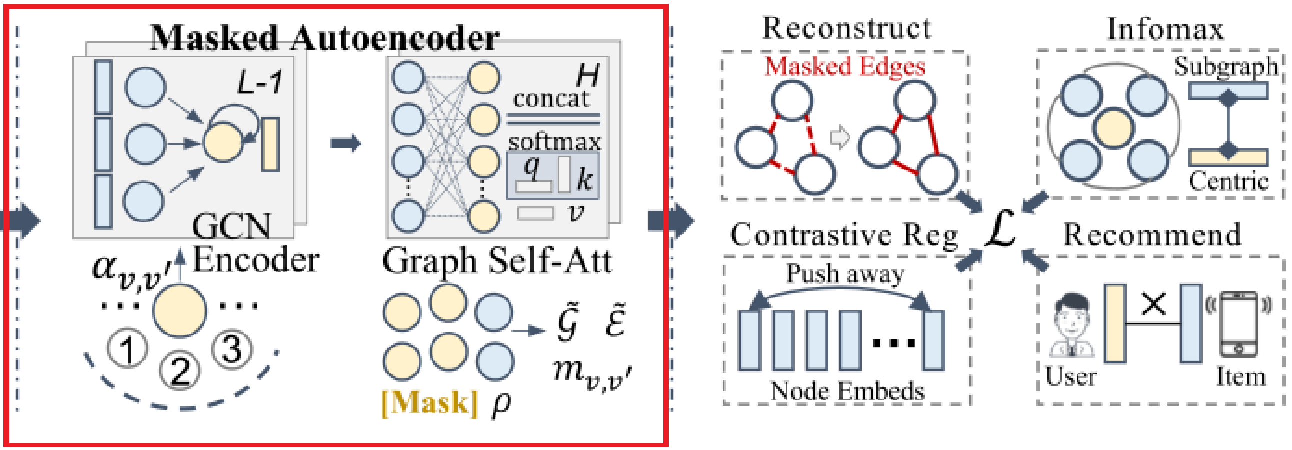
$$s_v = \psi(v; k) = \mathrm{sigm}(\mathbf{h}_v^\top \sum_{v' \in \mathcal{N}_v^k} \mathbf{h}_{v'} / (|\mathcal{N}_v^k| \cdot \|\mathbf{h}_v\| \cdot \|\mathbf{h}_{v'}\|)) \quad (4)$$

$$\psi'(v; k) = \log \psi(v; k) - \log(-\log(\mu)); \quad \mu \sim \mathrm{Uniform}(0, 1) \quad (5)$$

$$\mathcal{L}_{\text{InfoM}} = -\textstyle\sum_{v \in \mathcal{U} \cup \mathcal{I}} \psi(v; k).$$

Approach

$$\mathbf{h}_v^{l+1} = \alpha_{v,v}\cdot \mathbf{h}_v^l + \sum_{v'\in\mathcal{N}'_v} \alpha_{v,v'}\cdot \mathbf{h}_{v'}^l; \quad \alpha_{v,v'} = 1/\sqrt{|\mathcal{N}'_v||\mathcal{N}'_{v'}|} \tag{6}$$

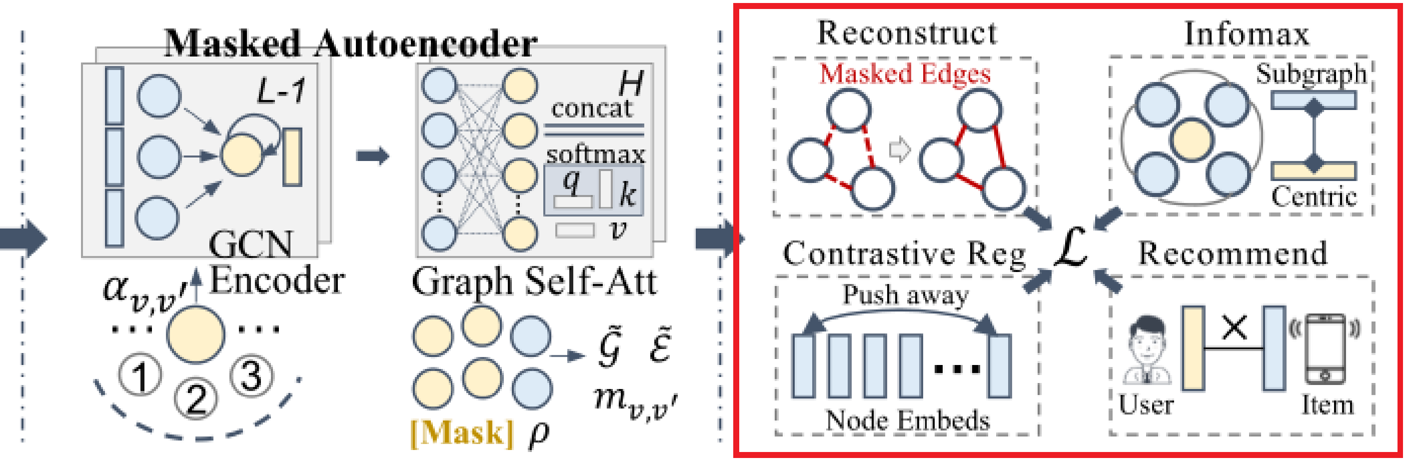


$$\tilde{\mathcal{G}} = \{\mathcal{U},\mathcal{V},\tilde{\mathcal{E}} = \bar{\mathcal{E}}\cup\mathcal{E}'\}; \quad \bar{\mathcal{E}} = \{(v_1,v_2)|v_1,v_2\in\bar{\mathcal{V}}\cup\tilde{\mathcal{V}}\}$$
$$\text{s.t. } |\bar{\mathcal{E}}| = |\mathcal{E}'|, \quad |\bar{\mathcal{V}}\cup\tilde{\mathcal{V}}| = \rho\cdot(|\mathcal{U}|+|\mathcal{I}|) \tag{7}$$

$$\mathbf{h}_v^{l+1} = \sum_{v'} \Big\Vert_{h=1}^{H} m_{v,v'}\beta^h_{v,v'}\mathbf{W}^h_{\mathrm{V}}\mathbf{h}^l_{v'}; \quad m_{v,v'} = \begin{cases} 1 & \text{if } (v,v')\in\tilde{\mathcal{E}} \\ 0 & \text{otherwise} \end{cases}$$
$$\beta^h_{v,v'} = \frac{\exp\bar{\beta}^h_{v,v'}}{\sum_{v'}\exp\bar{\beta}^h_{v,v'}}; \quad \bar{\beta}^h_{v,v'} = \frac{(\mathbf{W}^h_{\mathrm{Q}}\cdot\mathbf{h}^l_v)^\top\cdot(\mathbf{W}^h_{\mathrm{K}}\cdot\mathbf{h}^l_{v'})}{\sqrt{d/H}} \tag{8}$$

$$\mathcal{L}_{\text{recon}} = -\sum_{(v,v')\in\mathcal{E}\backslash\mathcal{E}'}\hat{\mathbf{h}}_v^\top\cdot\hat{\mathbf{h}}_{v'}; \quad \hat{\mathbf{h}}_v = \sum_{l=0}^{L}\mathbf{h}^l_v \tag{9}$$

Approach



$$\mathcal{L}_{\text{ssl}} = \sum_{u \in \mathcal{U}} \log \sum_{i \in \mathcal{I}} \exp \hat{\mathbf{h}}_u^\top \hat{\mathbf{h}}_i + \sum_{u \in \mathcal{U}} \log \sum_{u' \in \mathcal{U}} \exp \hat{\mathbf{h}}_u^\top \hat{\mathbf{h}}_{u'} + \sum_{i \in \mathcal{I}} \log \sum_{i' \in \mathcal{I}} \exp \hat{\mathbf{h}}_i^\top \hat{\mathbf{h}}_{i'} + \mathcal{L}_{\text{InfoM}} + \mathcal{L}_{\text{recon}} \quad (10)$$

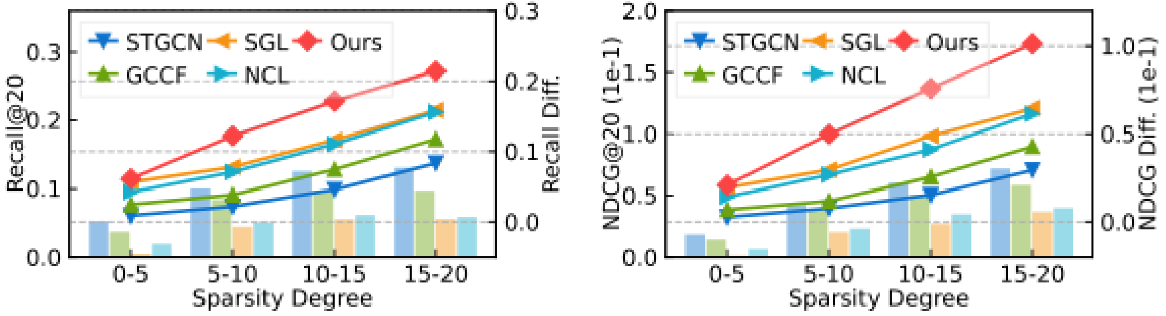
$$\mathcal{L} = -\sum_{(u,i) \in \mathcal{E}} \hat{\mathbf{h}}_u^\top \cdot \hat{\mathbf{h}}_i + \lambda_1 \cdot \mathcal{L}_{\text{ssl}} + \lambda_2 \cdot \|\Theta\|_{\text{F}}^2 \quad (11)$$

Experiment

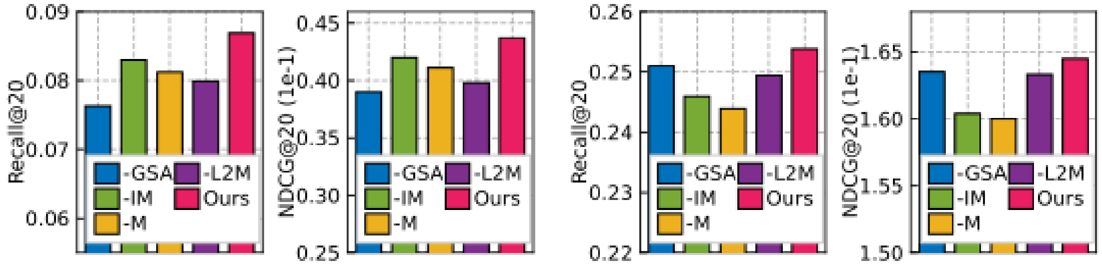
Table 2: Performance comparison on Gowalla, Yelp, and Amazon datasets in terms of *Recall* and *NDCG*.

Data	Metric	BiasMF	NCF	AutoR	PinSage	STGCN	GCMC	NGCF	GCCF	LightGCN	DGCF	SLRec	NCL	SGL	HCCF	*AutoCF*	p-val.
Gowalla	Recall@20	0.0867	0.1019	0.1477	0.1235	0.1574	0.1863	0.1757	0.2012	0.2230	0.2055	0.2001	0.2283	0.2332	0.2293	**0.2538**	$1.3e^{-10}$
	NDCG@20	0.0579	0.0674	0.0690	0.0809	0.1042	0.1151	0.1135	0.1282	0.1433	0.1312	0.1298	0.1478	0.1509	0.1482	**0.1645**	$4.9e^{-12}$
	Recall@40	0.1269	0.1563	0.2511	0.1882	0.2318	0.2627	0.2586	0.2903	0.3181	0.2929	0.2863	0.3232	0.3251	0.3258	**0.3441**	$9.3e^{-10}$
	NDCG@40	0.0695	0.0833	0.0985	0.0994	0.1252	0.1390	0.1367	0.1532	0.1670	0.1555	0.1540	0.1745	0.1780	0.1751	**0.1898**	$1.0e^{-9}$
Yelp	Recall@20	0.0198	0.0304	0.0491	0.0510	0.0562	0.0584	0.0681	0.0742	0.0761	0.0700	0.0665	0.0806	0.0803	0.0789	**0.0869**	$6.2e^{-7}$
	NDCG@20	0.0094	0.0143	0.0222	0.0245	0.0282	0.0280	0.0336	0.0365	0.0373	0.0347	0.0327	0.0402	0.0398	0.0391	**0.0437**	$1.0e^{-6}$
	Recall@40	0.0307	0.0487	0.0692	0.0743	0.0856	0.0891	0.1019	0.1151	0.1175	0.1072	0.1032	0.1230	0.1226	0.1210	**0.1273**	$1.2e^{-4}$
	NDCG@40	0.0120	0.0187	0.0268	0.0315	0.0355	0.0360	0.0419	0.0466	0.0474	0.0437	0.0418	0.0505	0.0502	0.0492	**0.0533**	$1.1e^{-5}$
Amazon	Recall@20	0.0324	0.0367	0.0525	0.0486	0.0583	0.0837	0.0551	0.0772	0.0868	0.0617	0.0742	0.0955	0.0874	0.0885	**0.1277**	$5.1e^{-13}$
	NDCG@20	0.0211	0.0234	0.0318	0.0317	0.0377	0.0579	0.0353	0.0501	0.0571	0.0372	0.0480	0.0623	0.5690	0.0578	**0.0879**	$8.0e^{-13}$
	Recall@40	0.0578	0.0600	0.0826	0.0773	0.0908	0.1196	0.0876	0.1175	0.1285	0.0912	0.1123	0.1409	0.1312	0.1335	**0.1782**	$7.3e^{-13}$
	NDCG@40	0.0293	0.0306	0.0415	0.0402	0.0478	0.0692	0.0454	0.0625	0.0697	0.0468	0.0598	0.0764	0.0704	0.0716	**0.1048**	$1.5e^{-13}$

Experiment



(a) Gowalla-Recall (b) Gowalla-NDCG



(a) Yelp Data (b) Gowalla Data

Figure 4: Module ablation study on Yelp and Gowalla.

-GSA. In this variant, a symmetry encoder-decoder

M (without the $\mathcal{L}_{\text{InfoM}}$ and $\mathcal{L}_{\text{recon}}$ regularization)

-IM removes the infomax-based loss $\mathcal{L}_{\text{InfoM}}$

masking strategy $\psi'(v;k)$ is replaced with the randomly edge masking in the variant **-L2M**.

Experiment

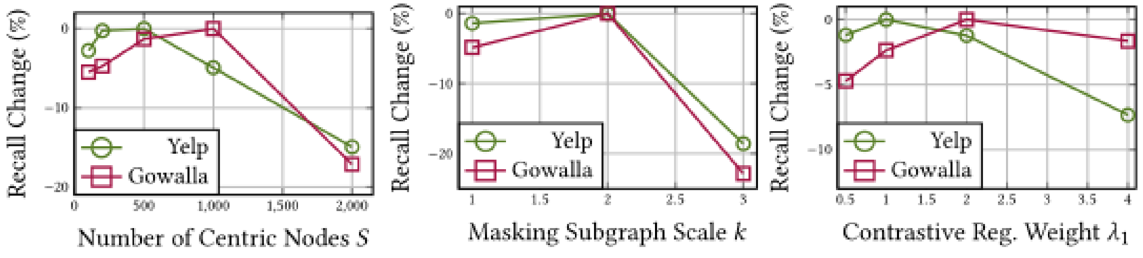


Figure 5: Hyperparameter study for the proposed AutoCF in terms of Recall@20 changes on Yelp and Gowalla datasets.

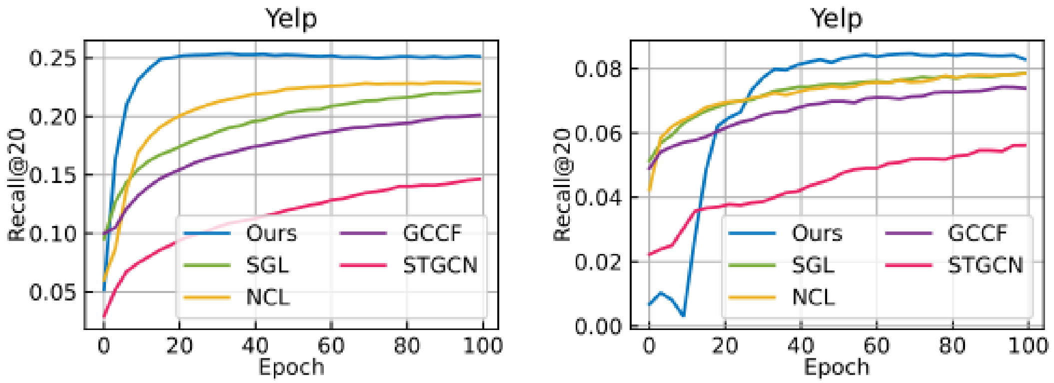
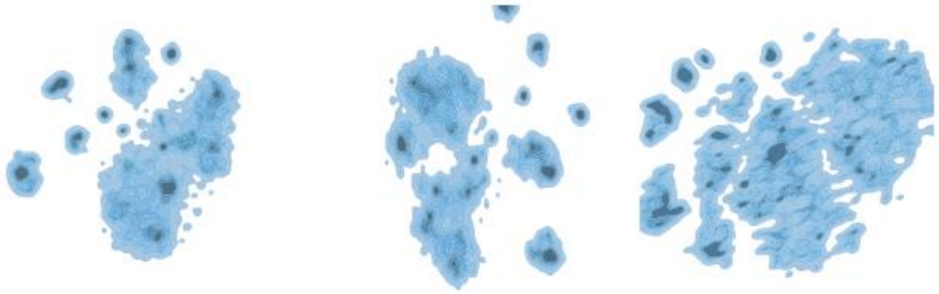


Figure 6: Convergence analysis *w.r.t* epochs while training.

Experiment

Table 3: Model efficiency study in terms of per-epoch training time (seconds) on Gowalla, Yelp, and Amazon datasets.

Model	DGCF	NCL	HCCF	SGL	Ours
Gowalla	12.03s	5.38s	6.00s	8.07s	7.09s
Yelp	11.47s	3.33s	4.07s	4.88s	4.06s
Amazon	85.54s	25.62s	48.28s	49.87s	59.81s



(a) HCCF (b) NCL (c) Ours

Figure 8: User embedding distribution of different methods with Gaussian kernel density estimation (KDE) on Yelp data.

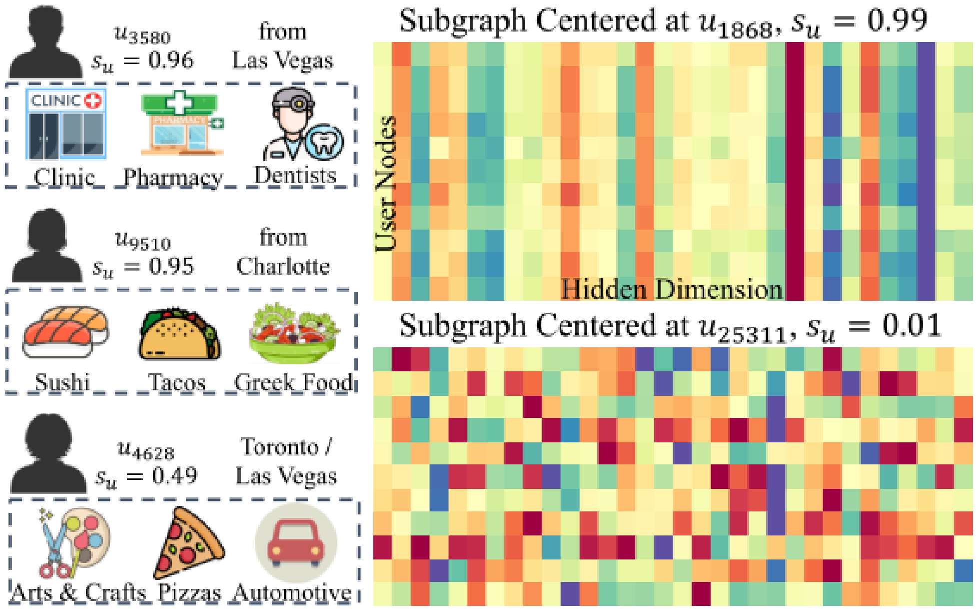


Figure 7: Case study for i) model interpretation in learning semantic relatednes of user interactions; ii) heatmaps of encoded user embeddings from two user-centric subgraphs.

Thank you!