Knowledge Graph Contrastive Learning for Recommendation

Yuhao Yang
University of Hong Kong
yuhao-yang@outlook.com

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

Lianghao Xia
University of Hong Kong
aka_xia@foxmail.com

Chenliang Li
Wuhan University
cllee@whu.edu.cn

——SIGIR 2022

https://github.com/yuh-yang/KGCL-SIGIR22

Reported by Yabo Yin

Artificial

gesis
Leibniz-Institut für Sozialwissenschaften

Introduction

1. Real-world knowledge graphs are often noisy and contain topic-irrelevant connections between items and entities.

2. The effectiveness of existing KG-aware recommendation methods largely relies on the high quality input knowledge graphs and are vulnerable to noise perturbation

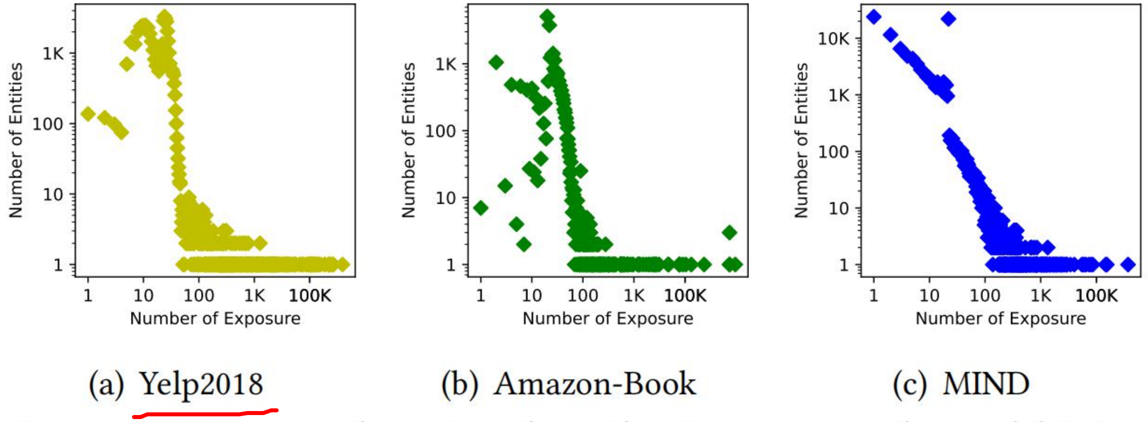


Figure 1: Long-tail entity distributions in real-world KGs.

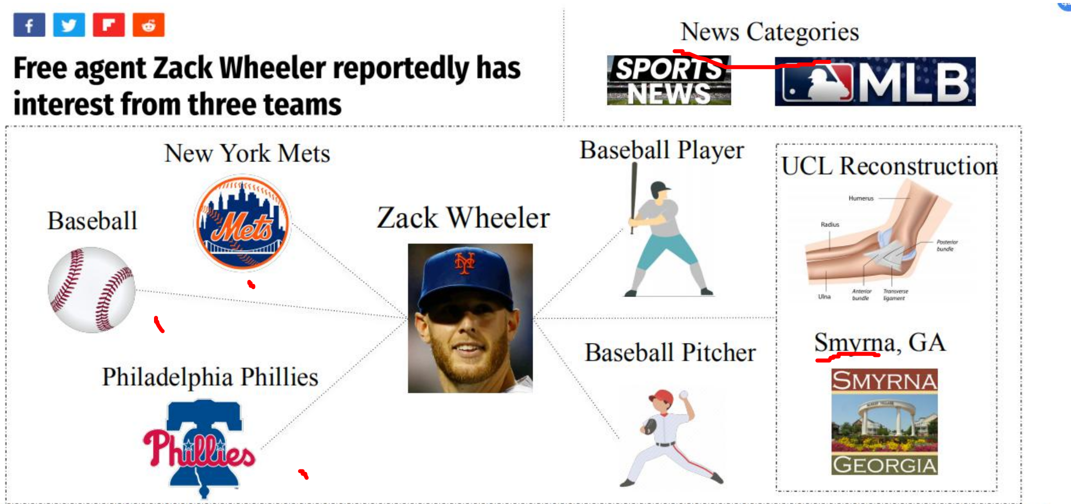


Figure 2: Illustrated example of news topic-irrelevant entities extracted from the knowledge graph of MIND dataset.

Method

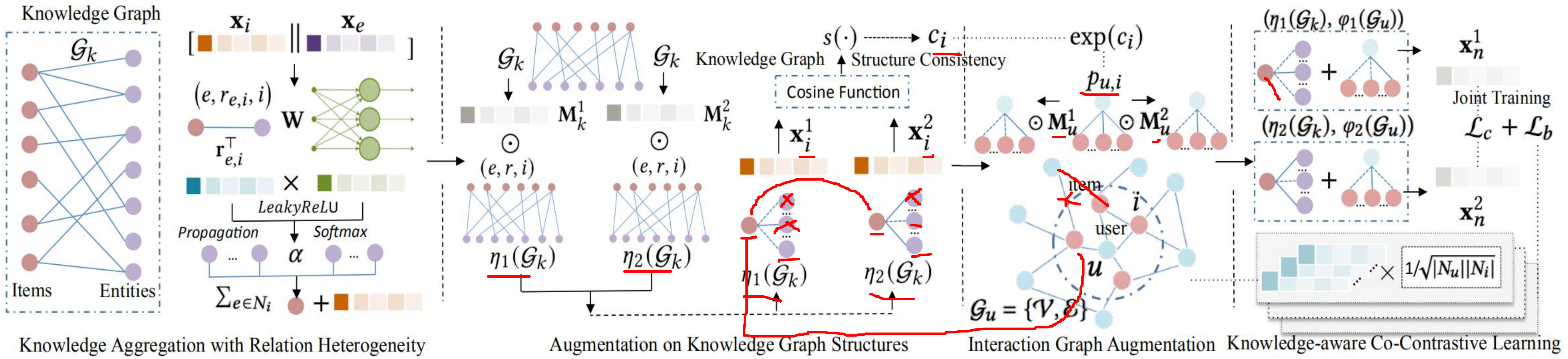
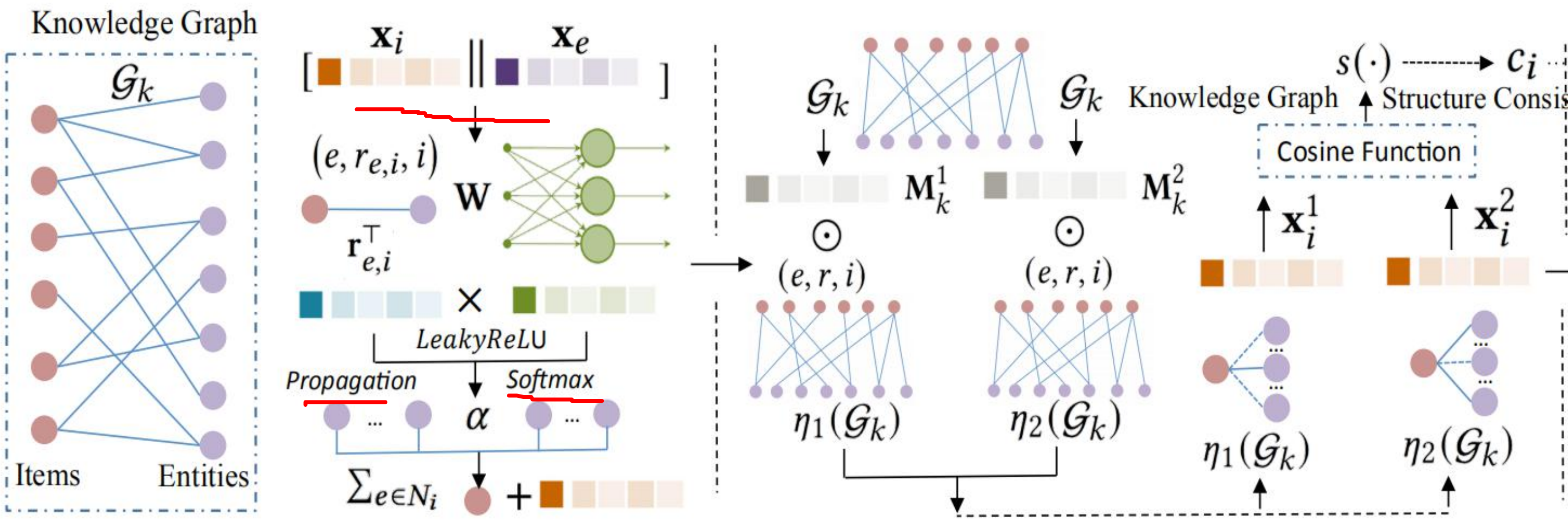


Figure 3: The overall architecture of our proposed KGCL. Knowledge-aware co-contrastive learning with augmentation functions on both knowledge graph $\eta(\cdot)$ and user-item interaction graph $\varphi(\cdot)$. Our contrastive objective $\mathcal{L}_c$ is jointly optimized with main embedding space shared by the knowledge graph aggregation and graph-based CF encoder.

Method



Knowledge Aggregation with Relation Heterogeneity

Augmentation on Knowledge Graph Structures

Problem Definition

user set $\mathcal{U}$ and an item set $\mathcal{I}$

user-item interaction matrix $\mathcal{Y} \in |\mathcal{U}| \times |\mathcal{I}|$

user-item interaction graph $\mathcal{G}_u = \{\mathcal{V}, \mathcal{E}\}$

$\mathcal{G}_k = \{(h, r, t)\}$ represent the knowledge graph

$$\mathbf{x}_i = \mathbf{x}_i + \sum_{e \in \mathcal{N}_i} \alpha\left(e, r_{e,i}, i\right) \mathbf{x}_e$$

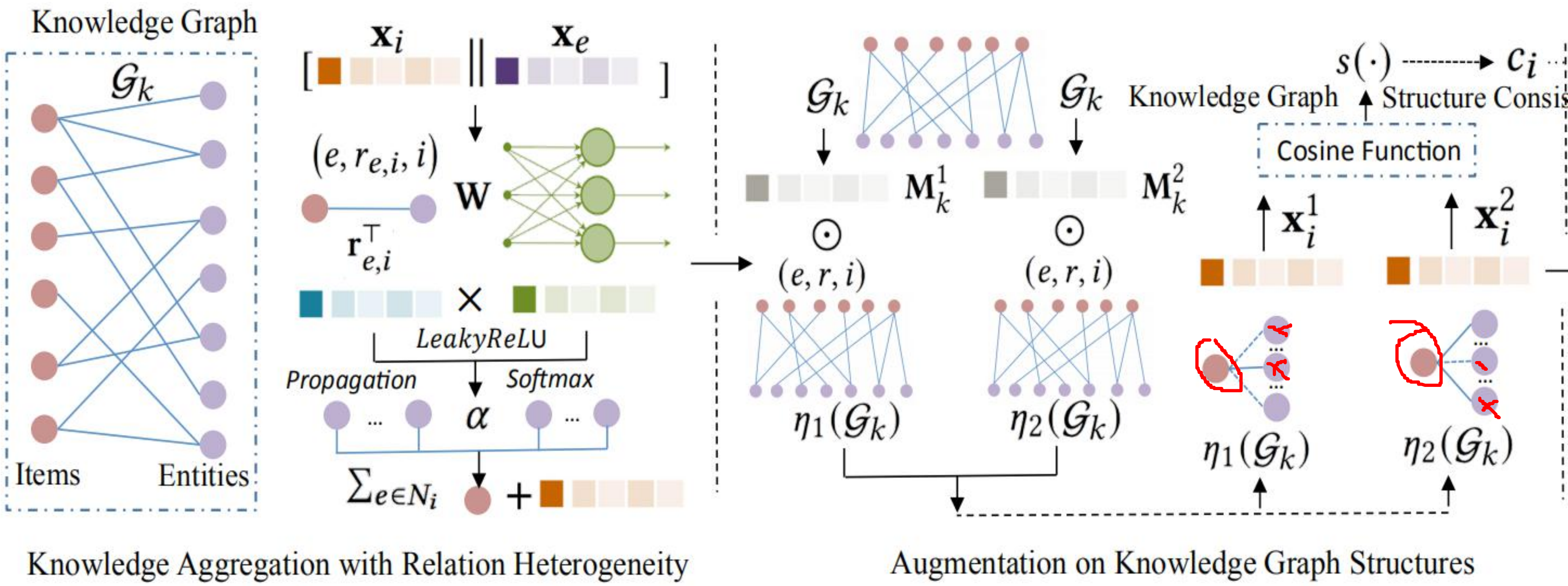
$$\alpha\left(e, r_{e,i}, i\right) = \frac{\exp\left(LeakyReLU\left(\mathrm{r}_{e,i}^{\top} \mathrm{W}\left[\mathbf{x}_e \| \mathbf{x}_i\right]\right)\right)}{\sum_{e \in \mathcal{N}_i} \exp\left(LeakyReLU\left(\mathrm{r}_{e,i}^{\top} \mathrm{W}\left[\mathbf{x}_e \| \mathbf{x}_i\right]\right)\right)} \qquad (1)$$

Semantic Representation Enhancement

$$\mathcal{L}_{TE} = \sum_{(h,r,t,t') \in \mathcal{G}_k} -\ln \sigma\left(f_d\left(\mathbf{x}_h, \mathbf{x}_r, \mathbf{x}_{t'}\right) - f_d\left(\mathbf{x}_h, \mathbf{x}_r, \mathbf{x}_t\right)\right) \qquad (2)$$

$$f_d = \left\|\mathbf{x}_h + \mathbf{x}_r - \mathbf{x}_t\right\|.$$

Method



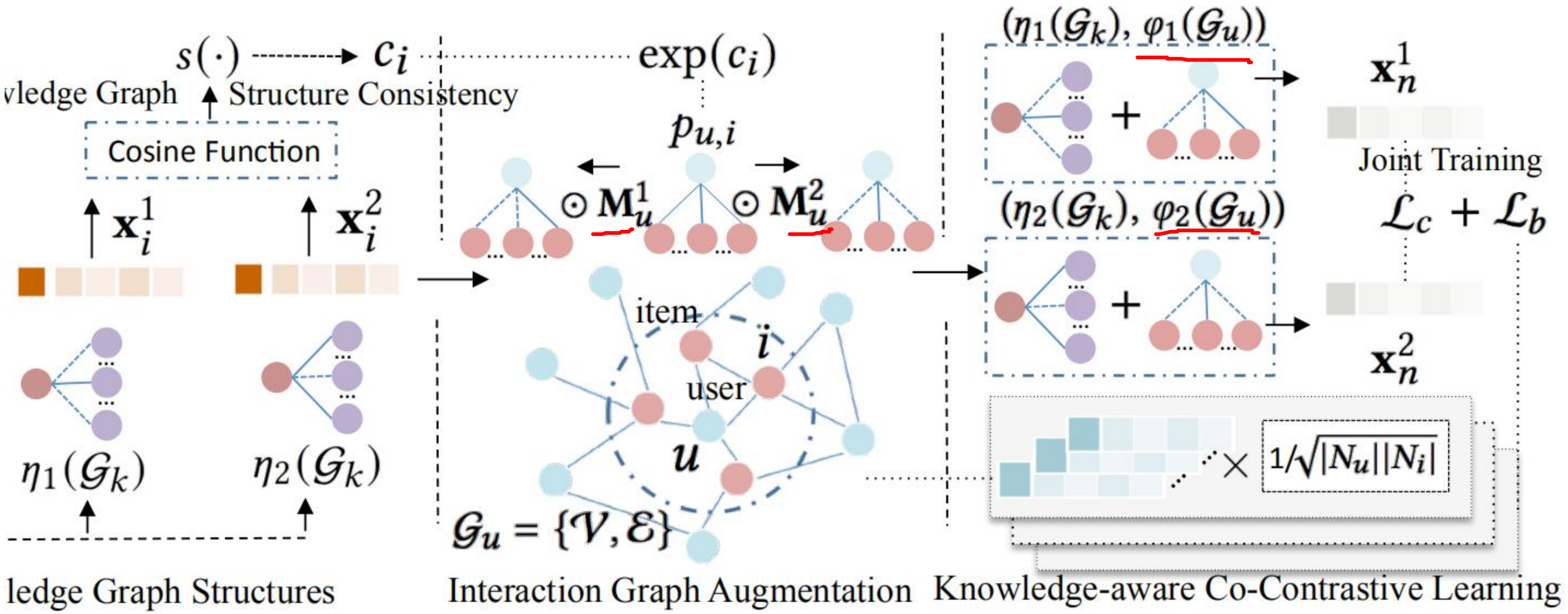
Augmentation on Knowledge Graph Structures.

$$\eta_1(\mathcal{G}_k) = ((e, r, i) \odot \mathbf{M}_k^1), \ \eta_2(\mathcal{G}_k) = ((e, r, i) \odot \mathbf{M}_k^2) \quad (3) \qquad (e, r, i) \in \mathcal{G}_k$$

$$c_i = s\left(f_k\left(\mathrm{x}_i, \eta_1(\mathcal{G}_k)\right), f_k\left(\mathrm{x}_i, \eta_2(\mathcal{G}_k)\right)\right) \quad (4)$$

f_k defined in Eq 1

Method



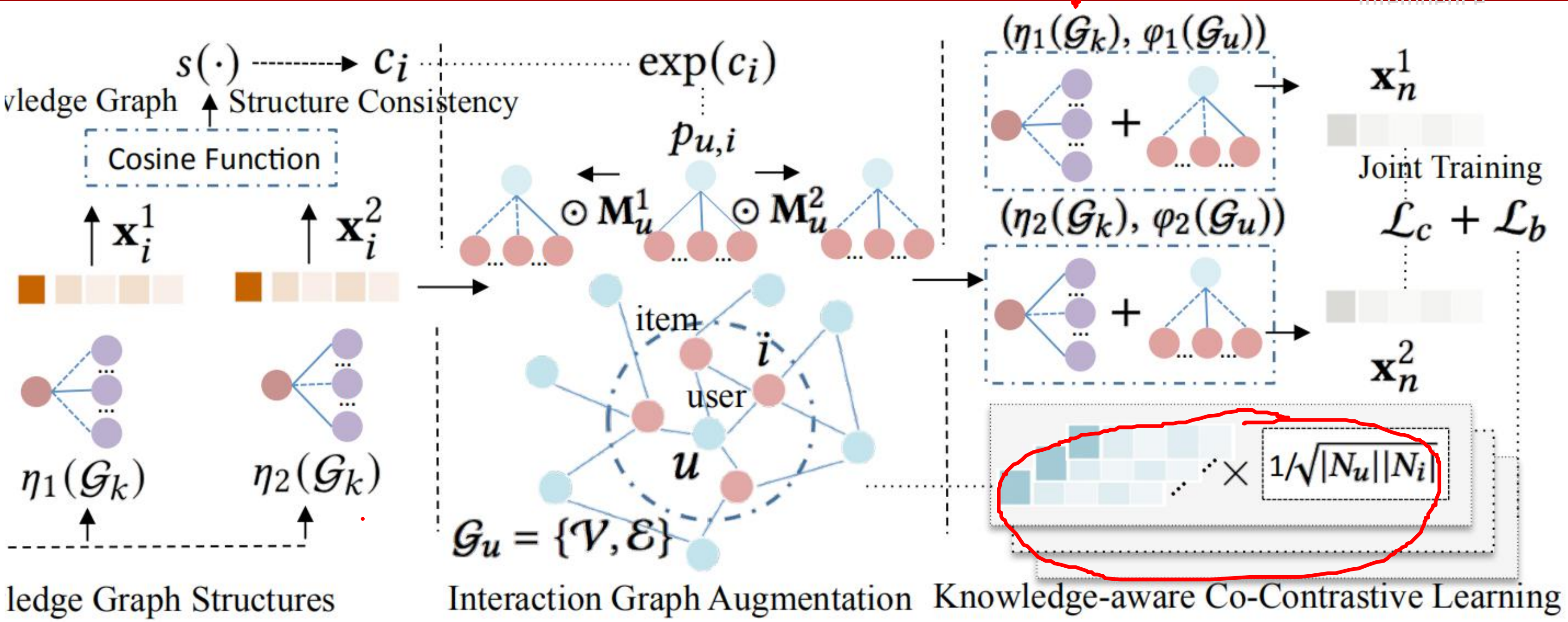
$$w_{u,i} = \exp(c_i);\ p'_{u,i} = \max\left(\frac{w_{u,i} - w^{min}}{w^{max} - w^{min}}, p_\tau\right) \tag{5}$$

$$p_{u,i} = p_a \cdot \mu_{p'} \cdot p'_{u,i}$$

sponding structure consistency score of c_i. We further perform the min-max normalization on $w_{u,i}$ with the truncation probability p_τ, to alleviate the low value effect. After that, the intermediate variable $p'_{u,i}$ is obtained and integrated with the mean value $\mu_{p'}$ to derive the value of dropout probability $p_{u,i}$. Here, p_a controls the strength of mean-based influence. With the probability $p_{u,i}$,

$$\varphi(\mathcal{G}_u) = (\mathcal{V}, \mathbf{M}_u^1 \odot \mathcal{E}),\ \varphi(\mathcal{G}_u) = (\mathcal{V}, \mathbf{M}_u^2 \odot \mathcal{E}) \tag{6}$$

Method



$$\mathbf{x}_u^{(l+1)} = \sum_{i \in N_u} \frac{\mathbf{x}_i^{(l)}}{\sqrt{|N_u||N_i|}};\ \mathbf{x}_i^{(l+1)} = \sum_{u \in N_i} \frac{\mathbf{x}_u^{(l)}}{\sqrt{|N_i||N_u|}} \tag{7}$$

$$\mathcal{L}_b = \sum_{u \in \mathcal{U}} \sum_{i \in N_u} \sum_{i' \notin N_u} -\log \sigma(\hat{y}_{u,i} - \hat{y}_{u,i'}) \tag{9}$$

$$\mathcal{L}_c = \sum_{n \in \mathcal{V}} -\log \frac{\exp(s(\mathbf{x}_n^1, \mathbf{x}_n^2)/\tau)}{\sum_{n' \in \mathcal{V}, n' \neq n} \exp(s(\mathbf{x}_n^1, \mathbf{x}_{n'}^2)/\tau)} \tag{8}$$

$$\mathcal{L} = \mathcal{L}_b + \lambda_1 \mathcal{L}_c + \lambda_2 \|\Theta\|_2^2, \tag{10}$$

Method

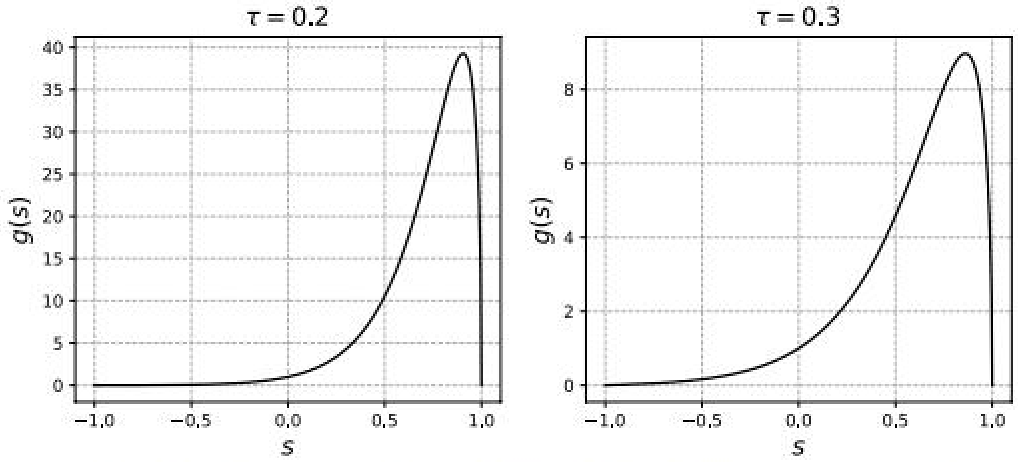


Figure 4: Distribution of gradient function $g(s)$ under $\tau = 0.2$ and $\tau = 0.3$. s is the similarity score between positive and negative instances. Hard negatives significant impact $g(s)$.

Theoretical Discussion of KGCL.

$$g(s) = \sqrt{1 - s^2} \exp\left(\frac{s}{\tau}\right) \tag{11}$$

$$\|\theta_f(U_f, v) - s_0\| < t_0 \tag{12}$$

$$\|\theta_r(U_f, v) - s_0\| \gg t_0 \tag{13}$$

$$\begin{aligned} \|\theta_r(U_f, v) - s_0\| \geq \|\theta_a(U_f, v) - s_0\| \gg \|\theta_f(U_f, v) - s_0\| \\ \|\theta_f(U_t, v) - s_0\| \gg \|\theta_a(U_t, v) - s_0\| \geq \|\theta_r(U_t, v) - s_0\| \end{aligned} \tag{14}$$

Experiments

Table 1: Statistics of experimented datasets.

Stats.	Yelp2018	Amazon-Book	MIND
# Users	45, 919	70, 679	300, 000
# Items	45, 538	24, 915	48, 957
# Interactions	1, 183, 610	846, 434	2, 545, 327
Density Degree	5.7×10^{-4}	4.8×10^{-4}	1.7×10^{-4}
	Knowledge Graph		
# Relations	42	39	90
# Entities	47, 472	29, 714	106, 500
# Triples	869, 603	686, 516	746, 270

Experiments

Table 2: Performance comparison of all methods on Yelp, Amazon and MIND. The superscript $*$ indicates the improvement is statistically significant where p-value < 0.01 level.

Model	Yelp2018		Amazon-book		MIND	
	Recall	NDCG	Recall	NDCG	Recall	NDCG
BPR	5.55%*	0.0375*	12.44%*	0.0658*	9.38%*	0.0469*
NCF	5.35%*	0.0346*	10.33%*	0.0532*	8.93%*	0.0436*
GC-MC	6.88%*	0.0453*	10.64%*	0.0534*	9.84%*	0.0491*
LightGCN	6.82%*	0.0443*	13.98%*	0.0736*	10.33%*	0.0520*
SGL	7.19%*	0.0475*	14.45%*	0.0766*	10.32%*	0.0539*
CKE	6.86%*	0.0431*	13.75%*	0.0685*	9.01%*	0.0382*
RippleNet	4.22%*	0.0251*	10.58%*	0.0549*	8.58%*	0.0407*
KGCN	5.32%*	0.0338*	11.11%*	0.0569*	8.87%*	0.0431*
KGAT	6.75%*	0.0432*	13.90%*	0.0739*	9.07%*	0.0442*
KGIN	7.12%*	0.0462*	14.36%*	0.0748*	10.44%*	0.0527*
CKAN	6.89%*	0.0441*	13.80%*	0.0726*	9.91%*	0.0499*
MVIN	6.91%*	0.0441*	13.98%*	0.0742*	9.62%*	0.0487*
KGCL	**7.56%**	**0.0493**	**14.96%**	**0.0793**	**10.73%**	**0.0551**

Experiments

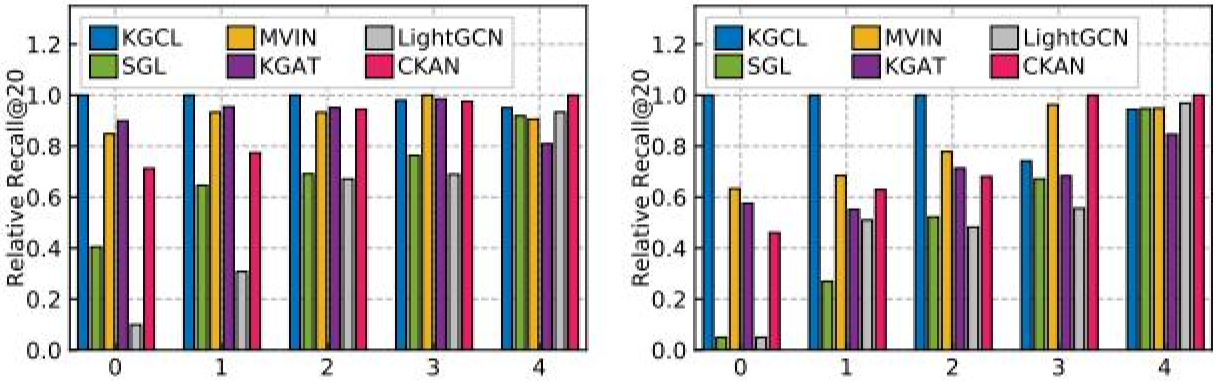
Table 3: Impact study of knowledge-aware graph augmentation schema with model variants of KGCL.

Model	Amazon-Book		MIND	
	Recall	NDCG	Recall	NDCG
KGCL	**14.96%**	**0.0793**	**10.73%**	**0.0551**
KGCL w/o KGA	14.85%	0.0788	10.57%	0.0546
KGCL w/o KGC	14.68%	0.0771	10.35%	0.0537

Table 4: Impact of τ and λ_1 on Amazon-Book dataset.

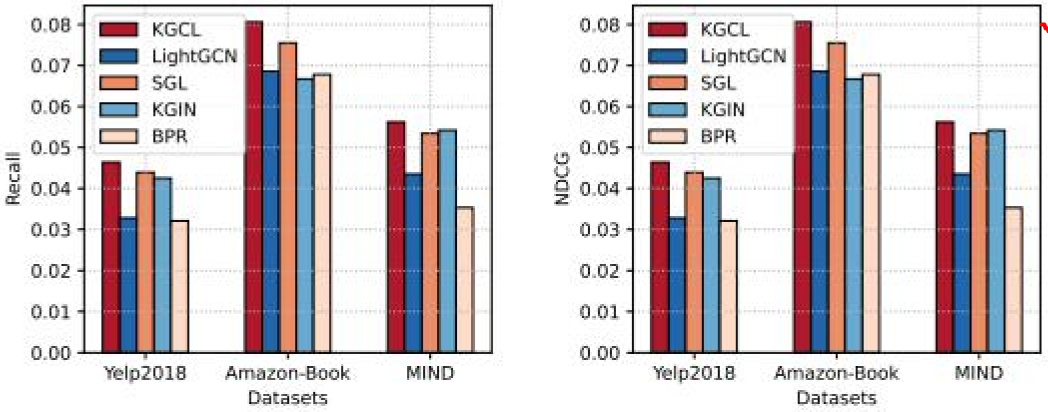
Metric	Recall@20				
λ_1, τ	0.1	0.2	0.3	0.4	0.5
10^{-1}	12.93%	**14.96%**	14.46%	13.94%	13.17%
10^{-2}	13.74%	13.68%	13.08%	12.39%	11.55%
10^{-3}	12.77%	11.94%	11.27%	10.62%	9.97%

Experiments



(a) Amazon-Book (b) Yelp2018

Figure 5: Performance with different interaction density degrees of items between KGCL and baselines. Recall values are normalized to range [0, 1] for better presentation.



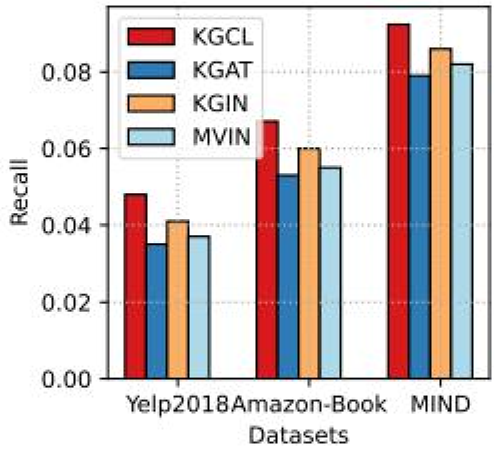
(a) Recall (b) NDCG

Figure 6: Comparison on cold-start users.

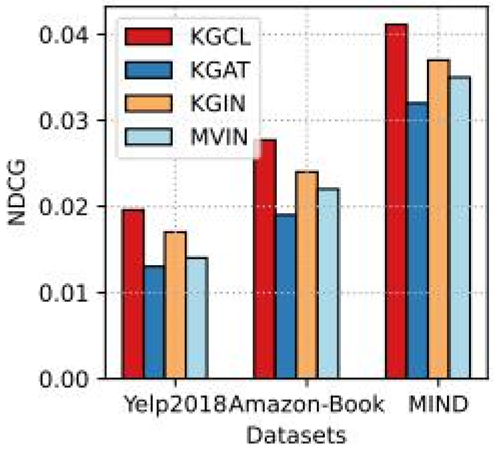
Experiments

Table 5: Performance in alleviating KG noise.

Model	Yelp2018		Amazon-book		MIND		Avg. Dec.
	Recall	NDCG	Recall	NDCG	Recall	NDCG	
KGAT	6.51%	0.0409	13.29%	0.0639	8.73%	0.0370	13.57%
KGIN	6.85%	0.0444	13.69%	0.0719	10.32%	0.0527	3.37%
MVIN	6.65%	0.0416	13.28%	0.0703	9.31%	0.0424	8.81%
KGCL	**7.52%**	**0.0490**	**14.93%**	**0.0787**	**10.69%**	**0.0550**	**0.58%**



(a) Recall



(b) NDCG

Figure 7: Recommendation performance comparison on items connected to long-tail KG entities.

Experiments

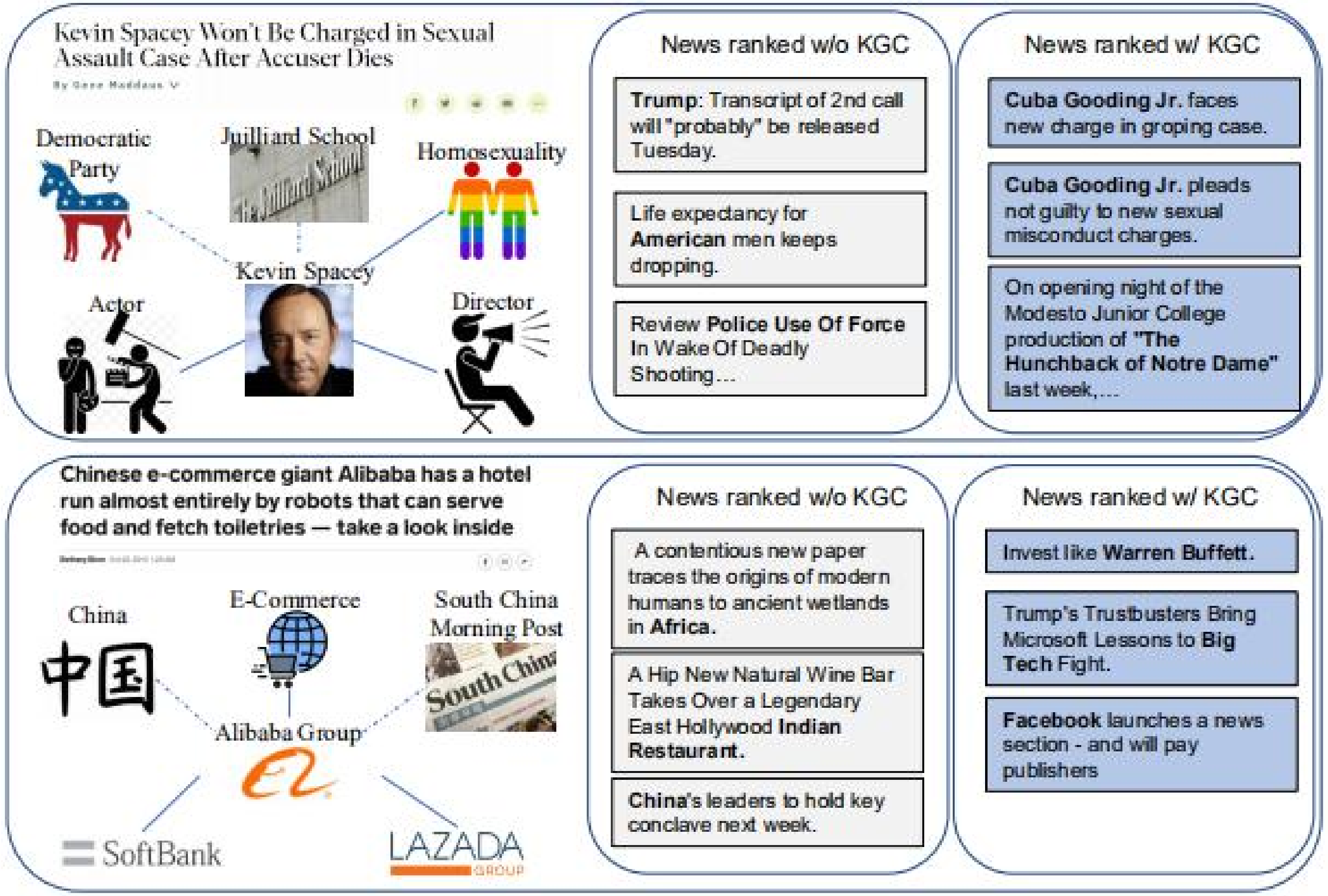


Figure 8: Two examples of relevant news ranked w/ and w/o KGCL. News in blue color means semantically relevant and bold font means the entities extracted from the news item.

Thank you!