南京航空航天大學
Nanjing University of Aeronautics and Astronautics

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

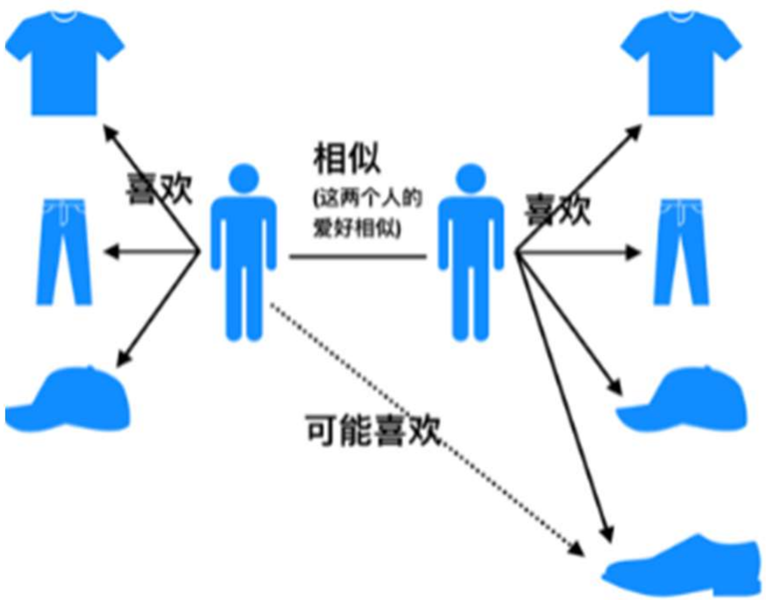
Zhiqiang Zhang
Ant Group
lingyao.zzq@antfin.com

Jun Zhou
Ant Group
jun.zhoujun@antfin.com

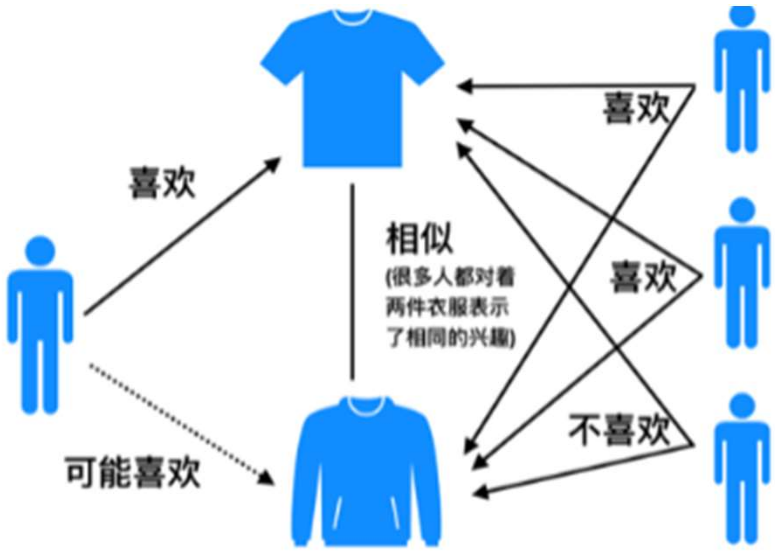
Meng Wang
Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology
Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
eric.mengwang@gmail.com

SIGIR 2023

Recommend System

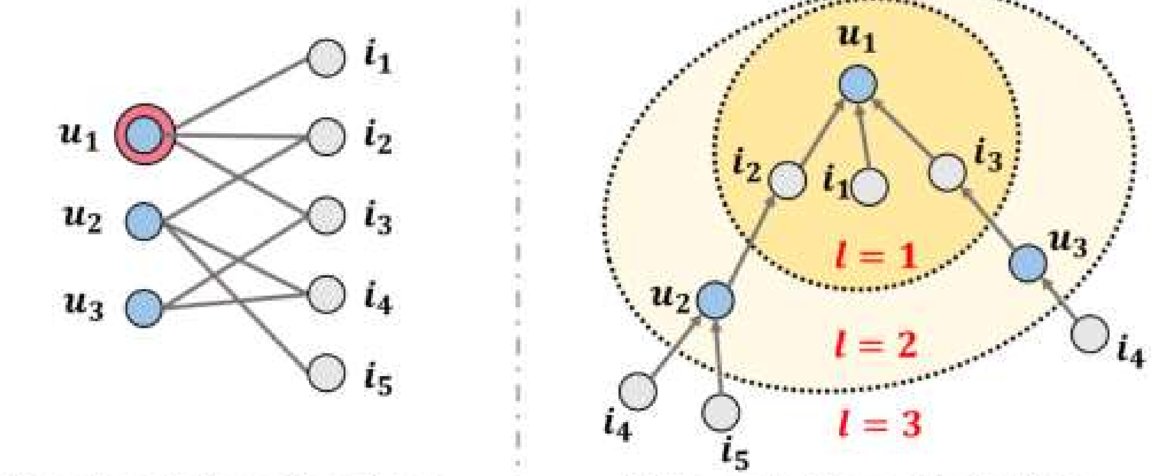


“人以群分”的基于用户的协同过滤



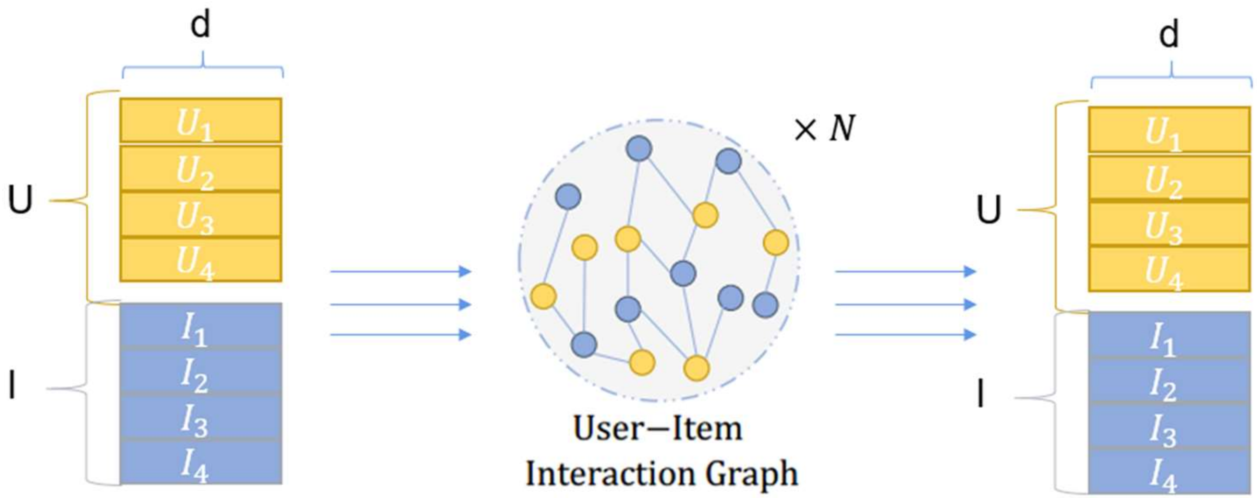
“物以类聚”的基于物品的协同过滤

GCF Methods



User-Item Interaction Graph

High-order Connectivity for u_1

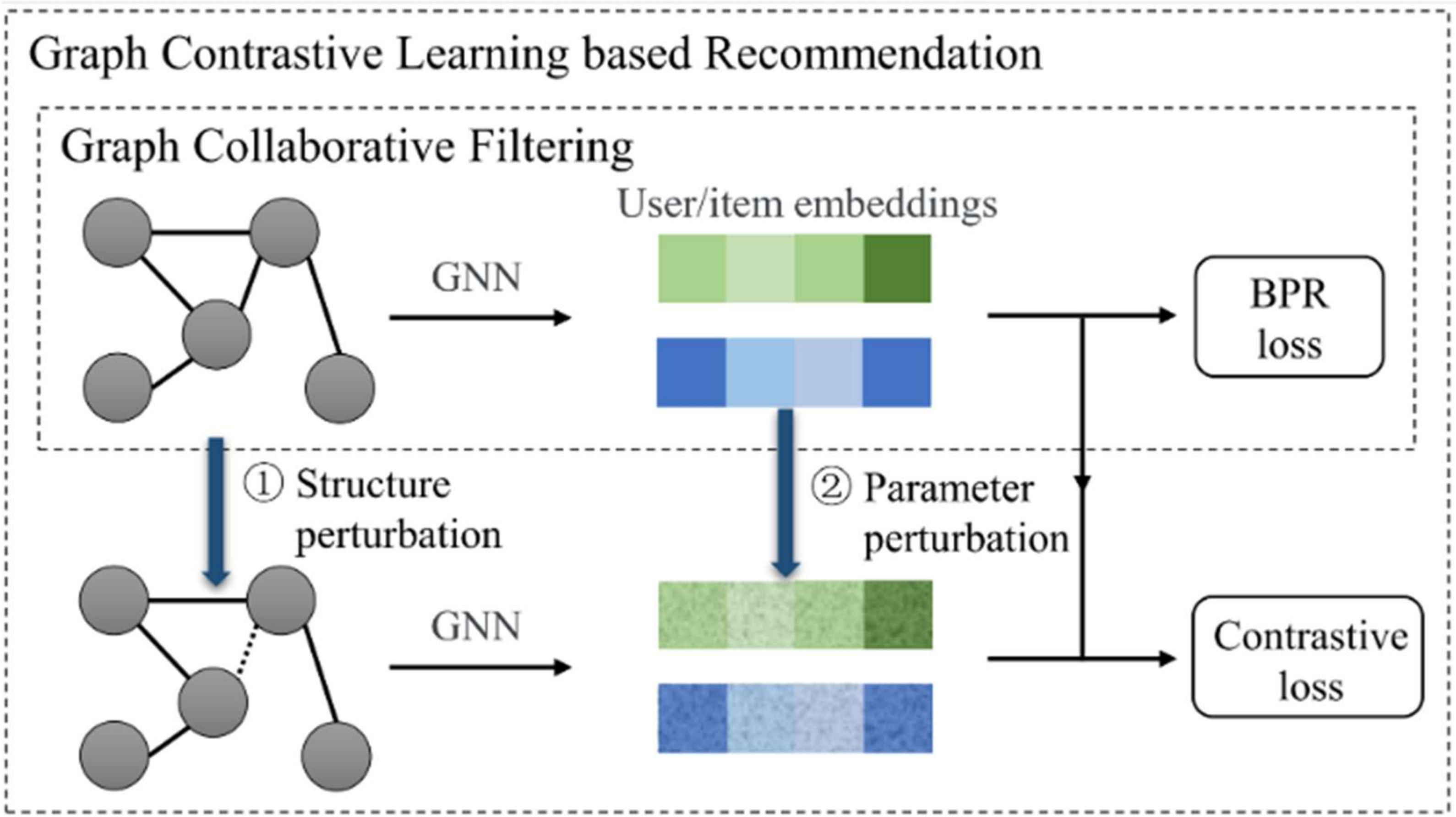


$$e_u^l = \sum_{i \in N_u} \frac{1}{\sqrt{|N_u||N_i|}} e_i^{l-1}$$

$$e_i^l = \sum_{i \in N_i} \frac{1}{\sqrt{|N_u||N_i|}} e_u^{l-1}$$

$$E^L = \hat{A} \cdot E^{L-1}$$

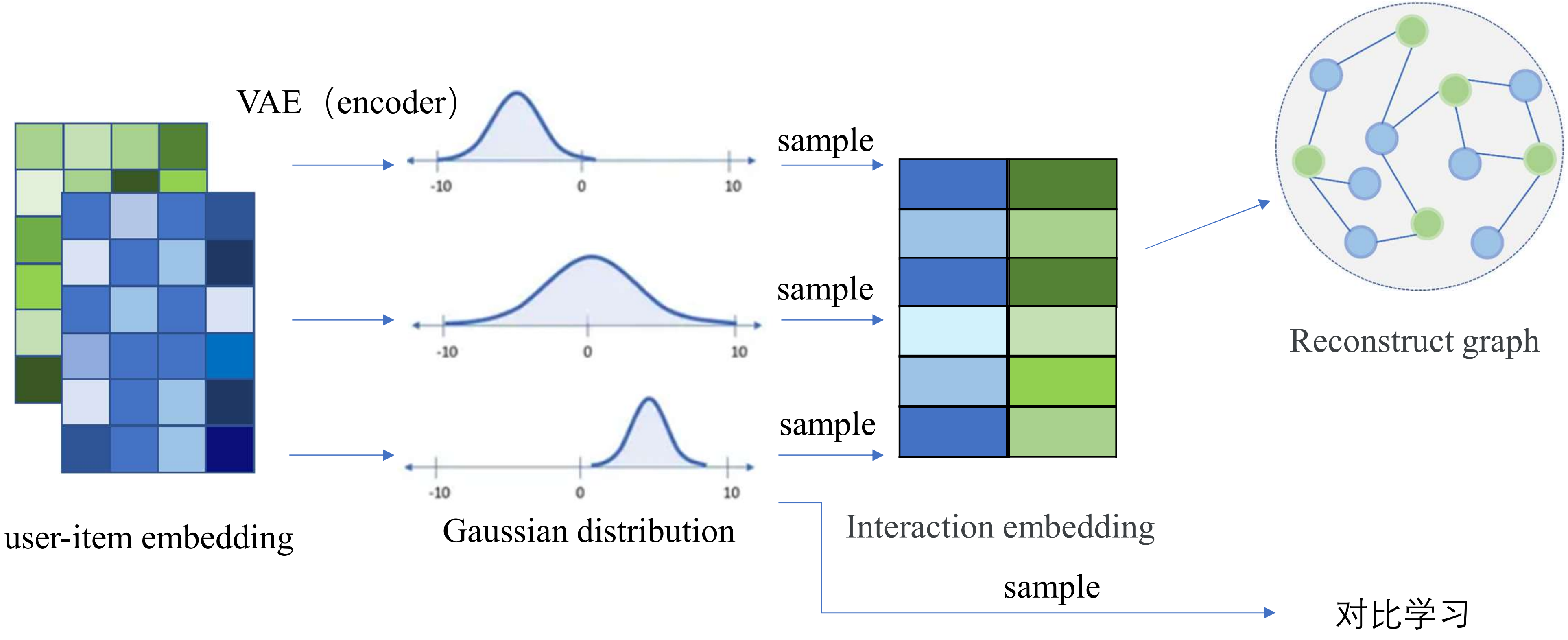
GCL Methods



$$L_{cl} = \sum_{i \in B} -\log \frac{\exp(e_i'^T e_i''/\tau)}{\sum_{j \in B} \exp(\frac{e_i'^T e_j''}{\tau})}$$

Abstract

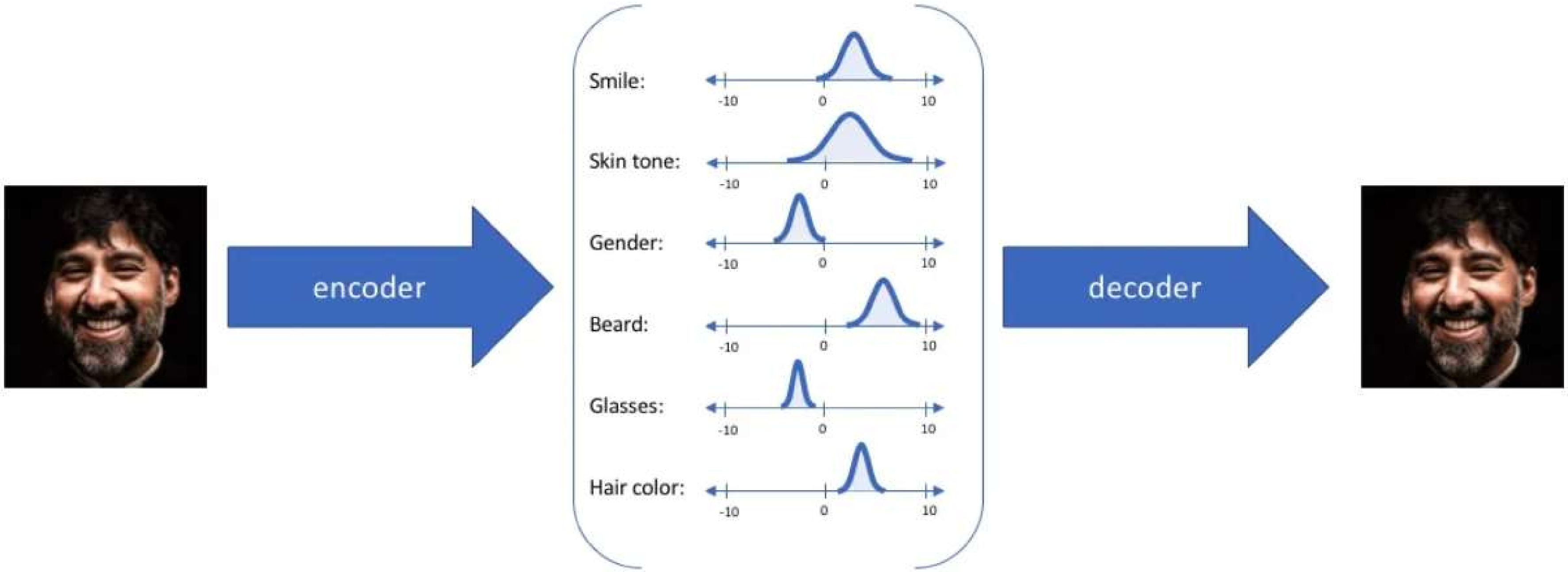
GCL现有方法缺陷:
1.基于图结构或emb的增强都可能破坏图结构，存在引入噪声的风险。
2.对于节点的特征扰动大小是随机的，忽略了图结点的自身特征。



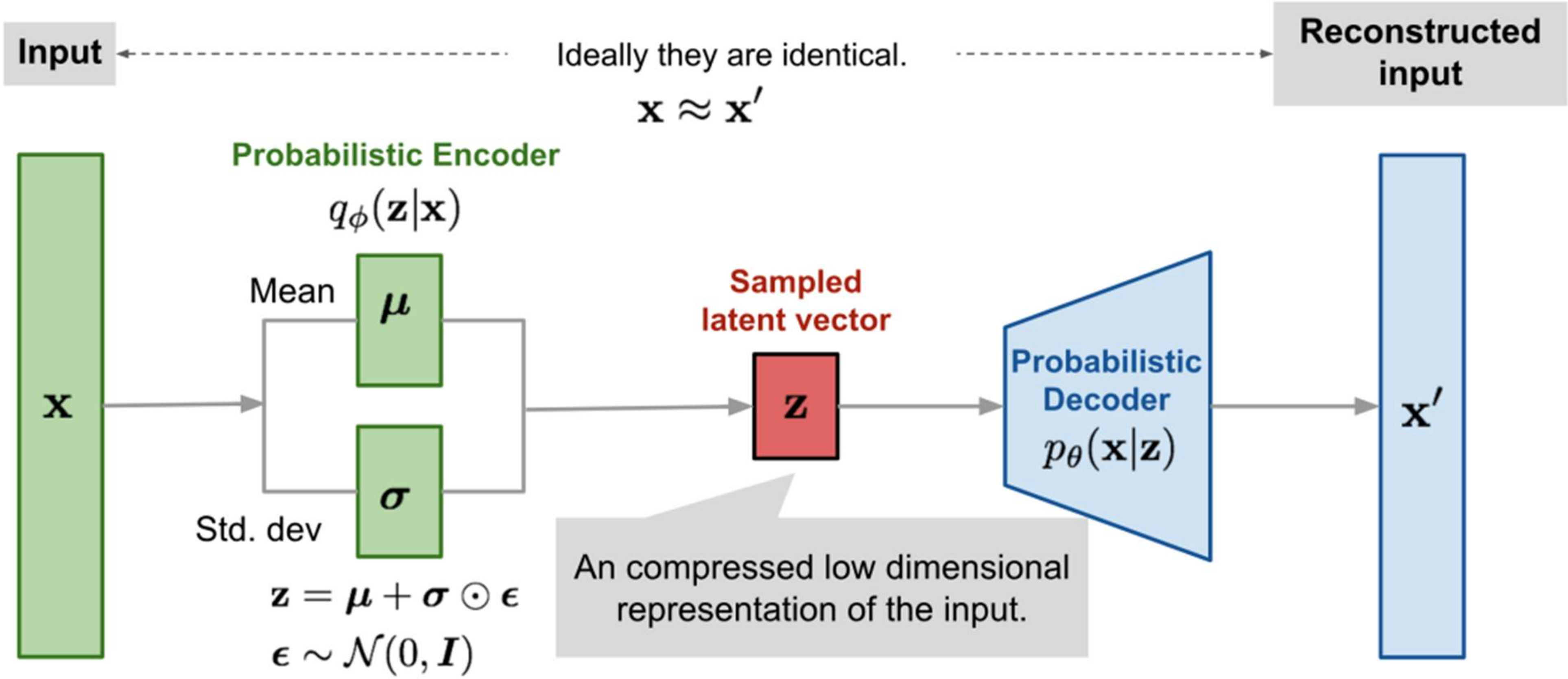
VAE(Variational Auto-Encoder)

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

VAE，又称变分自编码器，与传统自编码器是通过数值方式描述潜在空间的不同（例如GAN），VAE通过概率分布的方式表达目标空间



VAE(Variational Auto-Encoder)



VAE(Variational Auto-Encoder)

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Encoder（编码器）：

1. **目的**：使 $q_\phi(z|x)$近似后验分布 $p(z|x)$，得到较好的z 。
2. **实现**：通过一个神经网络，VAE定义了一个近似的后验分布 $q_\phi(z|x)$，通过高斯分布，预测其均值μ和标准差σ。
3. **优化目标**：通过最小化KL散度，使得 $q_\phi(z|x)$ 尽可能接近 $p(z|x)$ 。

Decoder（解码器）：

1. **目的**：重构输入数据x。
2. **实现**：从隐变量 z生成数据x′的$p_\theta(\mathrm{x}'|z)$
3. **优化目标**：通过最大化重构概率的对数，使得从 z重构的x′尽可能接近原始输入数据。

$$L_{\mathrm{EBLO}} = -E_{Z\sim q_\phi(z|x)}[lnp(x|z)] + KL(q_\phi(z|x)||p(z))$$

重构误差

衡量的是从隐变量 z重构输入数据 x的质量

KL散度

衡量后验概率$q_\phi(z|x)$与先验概率$p(z)$的差异

VAE(Variational Auto-Encoder)

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

ELBO(Evidence Lower Bound)证据下界

真实后验概率$p(z|x)$不可直接获取

通过$q_\phi(z|x)$近似$p(z|x)$

$$KL(q_\phi(z|x)||p(z|x))$$

$$= \int q_\phi(z|x) ln \frac{q_\phi(z|x)}{p(z|x)} dx$$

$$= E_{Z\sim q_\phi(z|x)}[ln \frac{q_\phi(z|x)}{p(z|x)}]$$

$$= E_{Z\sim q_\phi(z|x)}[lnq_\phi(z|x) - ln \frac{p(z|x)p(z)}{p(z)}]$$

$$= E_{Z\sim q_\phi(z|x)}\left[lnq_\phi(z|x) - \ln p(z) - lnp(x|z)\right] + lnp(x)$$

$$= KL(q_\phi(z|x)||p(z)) - E_{Z\sim q_\phi(z|x)}[lnp(x|z)] + lnp(x)$$

最大化对数似然函数　最小化$q_\phi(z|x)$和$p(z|x)$的距离

$$lnp(x) - KL(q_\phi(z|x)||p(z|x))$$

$$= E_{Z\sim q_\phi(z|x)}[lnp(x|z)] - KL(q_\phi(z|x)||p(z))$$

重构误差

缺乏KL散度方差会趋与0,
生成新样本能力下降

对数似然下界:

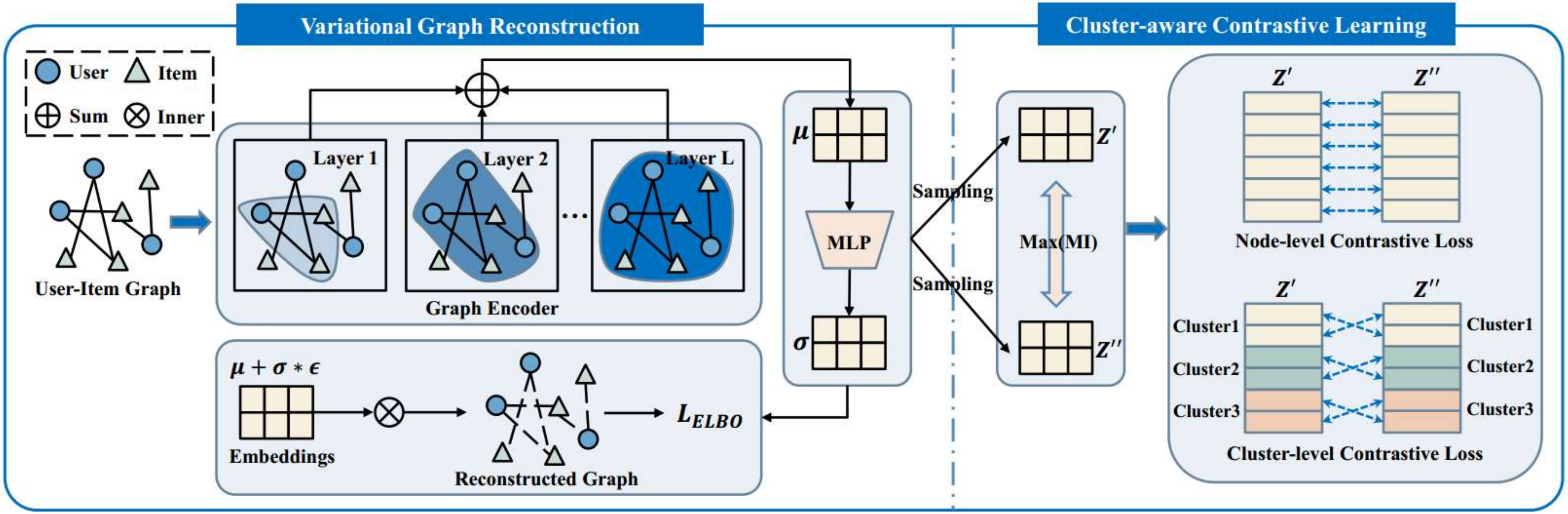
$$L_{ELBO} = -E_{Z\sim q_\phi(z|x)}[lnp(x|z)] + KL(q_\phi(z|x)||p(z))$$

$$lnp(x) = -L_{ELBO} + KL(q_\phi(z|x)||p(z|x))$$

当最小化L_{ELBO}时间接最小化$KL(q_\phi(z|x)||p(z|x))$
使$q_\phi(z|x)$近似真实后验概率$p(z|x)$

METHODOLOGY

南京航空航天大學
Nanjing University of Aeronautics and Astronautics



Variational Graph Reconstruction

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

为了尽可能的保留原始图信息和结构，本文推断每个节点的高斯分布进行采样，重构输入图或生成对比视图。

为了实现对图的重构，我们需要学习隐向量分布 Z

$$\hat{A} \sim p_\theta(A|Z)$$

根据VAE：

$$q_\phi(Z|A, E^0) = \prod_{I=0}^{M+N-1} q_\phi(z_i|A, E^0)$$

$$= \prod_{I=0}^{M+N-1} N\left(z_i \middle| \left(\mu_\phi(i)\right), diag(\sigma_\phi^2(i))\right)$$

→ $p_\theta(A|Z)$

Variational Graph Reconstruction

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

$$\mu_i^l = \sum_{j \in N_i} \frac{1}{\sqrt{|N_i||N_j|}} \mu_i^{l-1}, \mu^0 = E^0$$

$$\mu_i = \frac{1}{L} \sum_{j \in N_i} \mu_i^l$$

$$\sigma = MLP(\mu) = \exp(\mu W + b), W \in R^{d \times d}, b \in R^d$$

$$z_i = \mu_i + \sigma_i \cdot \varepsilon, \varepsilon \sim \mathrm{N}(0,1)$$ 作为节点的embedding

$$p(\hat{A}|Z) = \prod_{i=0}^{M+N-1} \prod_{j=0}^{M+N-1} p(\hat{A}_{i,j}|z_i, z_j) = \prod_{i=0}^{M+N-1} \prod_{j=0}^{M+N-1} sigmod(z_i^T z_j)$$

$$L_{ELBO} = -E_{Z \sim q_\phi(Z|\hat{A}, E^0)}[\log\left(p_\theta(\hat{A}|Z)\right) + KL[q_\phi(Z|\hat{A}, E^0)||p(Z)]]$$

$$\approx \sum_{i=0}^{M-1} \sum_{(j^+, j^-) \in D_i} \log \mathrm{sigmod}(\hat{r}_{i,j^+} - \hat{r}_{i,j^-}) + KL[q_\phi(Z|\hat{A}, E^0)||p(Z)]]$$

bpr损失验证重构图的准确性　确保q_ϕ的分布趋于正态分布

Contrastive Learning

$$z_i = \mu_i + \sigma_i \cdot \varepsilon, \varepsilon \sim \mathrm{N}(0,1)$$

对分布进行采样生成对比视图

$$z_i' = \mu_i + \sigma_i \cdot \varepsilon', \varepsilon' \sim \mathrm{N}(0,1)$$

$$z_i'' = \mu_i + \sigma_i \cdot \varepsilon'', \varepsilon'' \sim \mathrm{N}(0,1)$$

Node-level Contrastive Loss

$$L_N^U = \sum_{i \in B_u} -\log \frac{\exp(z_i'^T z_i''/\tau)}{\sum_{j \in B_u} \exp(\frac{z_i'^T z_j''}{\tau})}$$

$$L_N = L_N^U + L_N^I$$

Cluster-level Contrastive Loss

对节点进行聚类，不同集群的节点对一致性最小，相同集群的节点对一致性最大。

$$p(i,j) = \sum_{k=0}^{K_u - 1} p(c_k^u | z_i) p(c_k^u | z_j)$$

$$L_C^U = \sum_{i \in B_u} -\frac{1}{SP(i)} \log \frac{\sum_{j \in B_u, j!=i} p(i,j) \exp(z_i'^T z_\mathrm{j}''/\tau)}{\sum_{j \in B_u, j!=i} \exp(\frac{z_i'^T z_j''}{\tau})}$$

$$SP(i) = \sum_{j \in B_u, j!=i} p(i,j)$$

$$L_{cl} = \mathrm{L_N} + \gamma L_C$$

Algorithm

Algorithm 1: The Algorithm of *VGCL*

Input: user-item bipartite graph $\mathcal{G}$;

Output: Parameters $\Theta_{GNN} = \mathbf{E}^0$ and $\Theta_{MLP} = [\mathbf{W}, \mathbf{b}]$;

Randomly initialize parameters Θ_{GNN} and Θ_{MLP} ;
while not converged **do**
    Sample a batch of training data;
    Calculate graph inference parameters μ and σ (Eq.(12) to Eq.(13));
    Estimate node distribution $\mathbf{Z}$ by parameterization (Eq.(14));
    Generate contrastive instances $\mathbf{Z}'$ and $\mathbf{Z}''$ by multiple samplings (Eq.(17), Eq.(18));
    Compute prototypes $\mathbf{C}_u$ and $\mathbf{C}_v$ based on K-Means clustering algorithm;
    Compute node-level contrastive loss $\mathcal{L}_N$ (Eq.(19));
    Compute cluster-level contrastive loss $\mathcal{L}_C$ (Eq.(23), Eq.(24));
    Compute variational graph reconstruction loss $\mathcal{L}_{ELBO}$ (Eq.(26));
    Update all parameters according to (Eq.(28));
end while
Return $\Theta_{GNN} = \mathbf{E}^0$ and $\Theta_{MLP} = [\mathbf{W}, \mathbf{b}]$.

$$\mu_i = \frac{1}{L} \sum_{j \in N_i} \sum_{j \in N_i} \frac{1}{\sqrt{|N_i||N_j|}} \mu_i^{l-1}$$

$$\sigma = MLP(\mu) = \exp(\mu W + b)\,, W \in R^{d \times d}, b \in R^d$$

$$z_i = \mu_i + \sigma_i \cdot \varepsilon, \varepsilon \sim \mathrm{N}(0,1)$$

$$z_i' = \mu_i + \sigma_i \cdot \varepsilon', \varepsilon' \sim \mathrm{N}(0,1) \quad z_i'' = \mu_i + \sigma_i \cdot \varepsilon'', \varepsilon'' \sim \mathrm{N}(0,1)$$

重构交互图$\hat{A}$，评估z的质量

EXPERIMENTS

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.

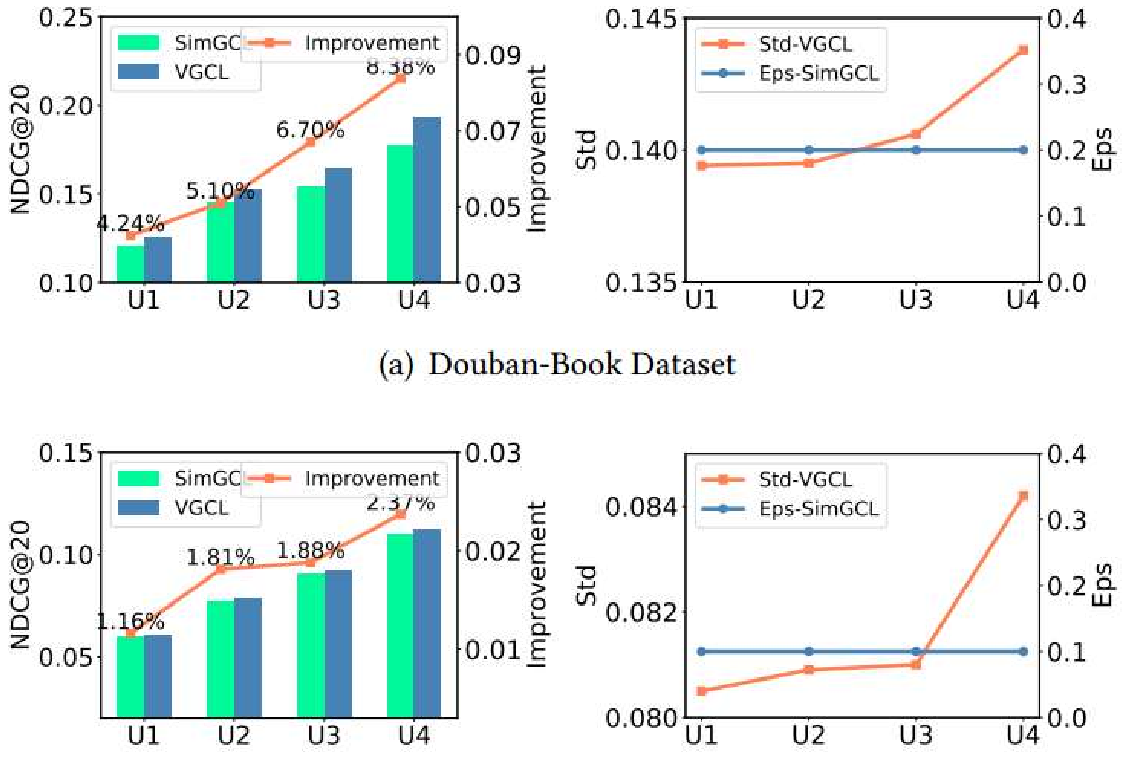
Models	Douban-Book				Dianping				Movielens-25M			
	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20
BPR-MF	0.0869	0.0949	0.1296	0.1045	0.0572	0.0443	0.0934	0.0557	0.2152	0.2011	0.3163	0.2343
LightGCN	0.1042	0.1195	0.1516	0.1278	0.0679	0.0536	0.1076	0.0660	0.2258	0.2192	0.3263	0.2509
Multi-VAE	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	0.1218	0.1470	0.1731	0.1540	0.0768	0.0606	0.1208	0.0743	0.2428	0.2356	0.3491	0.2690
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**

EXPERIMENTS

Table 3: Ablation study of VGCL, VGCL-w/o C ... thout the variational graph reconstruction part.

Models	Douban-Book	
	R@20	N@2(
LightGCN	0.1512(-)	0.1271(
SimGCL	0.1731(14.48%)	0.1540(+21
VGCL-w/o C	0.1776(+17.46%)	0.1575(+23
VGCL-w/o V	0.1722(+13.89%)	0.1547(+21
VGCL	**0.1829(+20.97%)**	**0.1638(+28**

) / 21%) / 31%) / 59%) / **51%)**



(a) Douban-Book Dataset

(b) Dianping Dataset

Figure 3: Performance comparisons under different user groups.

Thanks

VAE(Variational Auto-Encoder)

南京航空航天大學 Nanjing University of Aeronautics and Astronautics

ELBO(Evidence Lower Bound)证据下界

变分推断中采用KL散度作为度量两个概率分布的相似程度，这里用它作为 $q_\theta(z|X)$ 和 $p(z|X)$ 之间的“距离”： x

$$
\begin{aligned}
&KL(q_\theta(z|x)||p(z|x)) \\
&= \int q_\theta(z|x) \ln \frac{q_\theta(z|x)}{p(z|x)} \mathrm{d}z \\
&= \mathbb{E}_{z\sim q_\theta(z|x)}[\ln \frac{q_\theta(z|x)}{p(z|x)}] \\
&= \mathbb{E}_{z\sim q_\theta(z|x)}[\ln q_\theta(z|x) - \ln p(z|x)] \\
&= \mathbb{E}_{z\sim q_\theta(z|x)}[\ln q_\theta(z|x) - \ln \frac{p(x|z)p(z)}{p(x)}] \\
&= \mathbb{E}_{z\sim q_\theta(z|x)}[\ln q_\theta(z|x) - \ln p(z) - \ln p(x|z)] + \ln p(x) \\
&= KL(q_\theta(z|x)||p(z)) - \mathbb{E}_{z\sim q_\theta(z|x)}[\ln p(x|z)] + \ln p(x)
\end{aligned}
$$

整理后得到

$$
\ln p(x) - KL(q_\theta(z|x)||p(z|x)) = \mathbb{E}_{z\sim q_\theta(z|x)}[\ln p(x|z)] - KL(q_\theta(z|x)||p(z))
$$

Variational Graph Reconstruction

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

输入: $\mathcal{g} = \{U \cup V, A\}, E^0$

目的: learn distributions Z

如何评估生成的Z的好坏?

$$\mu_i^l = \sum_{j \in N_i} \frac{1}{\sqrt{|N_i||N_j|}} \mu_i^{l-1}$$

$$\mu_i = \frac{1}{L} \sum_{j \in N_i} \mu_i^l$$

$$\sigma = MLP(\mu) = \exp(\mu W + b), W \in R^{d \times d}, b \in R^d$$

$$z_i = \mu_i + \sigma_i \cdot \varepsilon, \varepsilon \sim \mathrm{N}(0,1)$$

图重构:

$$\hat{A} \sim \mathrm{p}_\theta(A|Z)$$

$$q_\phi(Z|A, E^0) = \prod_{I=0}^{M+N-1} q_\phi(z_i|A, E^0)$$

$$= \prod_{I=0}^{M+N-1} N\left(z_i \middle| \left(\mu_\phi(i)\right), diag\left(\sigma_\phi^2(i)\right)\right)$$

→ $p_\theta(A|Z)$