



DisenKGAT: Knowledge Graph Embedding with Disentangled Graph Attention Network

汇报人：陈丽行、许越玥、周俊池



目录 | CONTENT

- 1 相关概念
- 2 研究背景
- 3 模型介绍
- 4 实验分析

相关概念

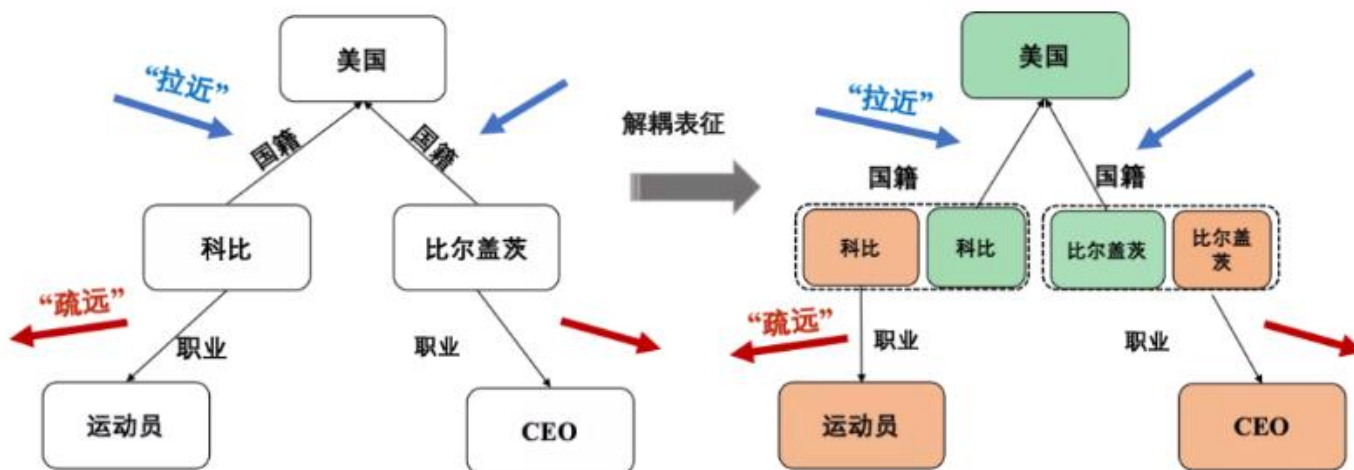


知识
图谱

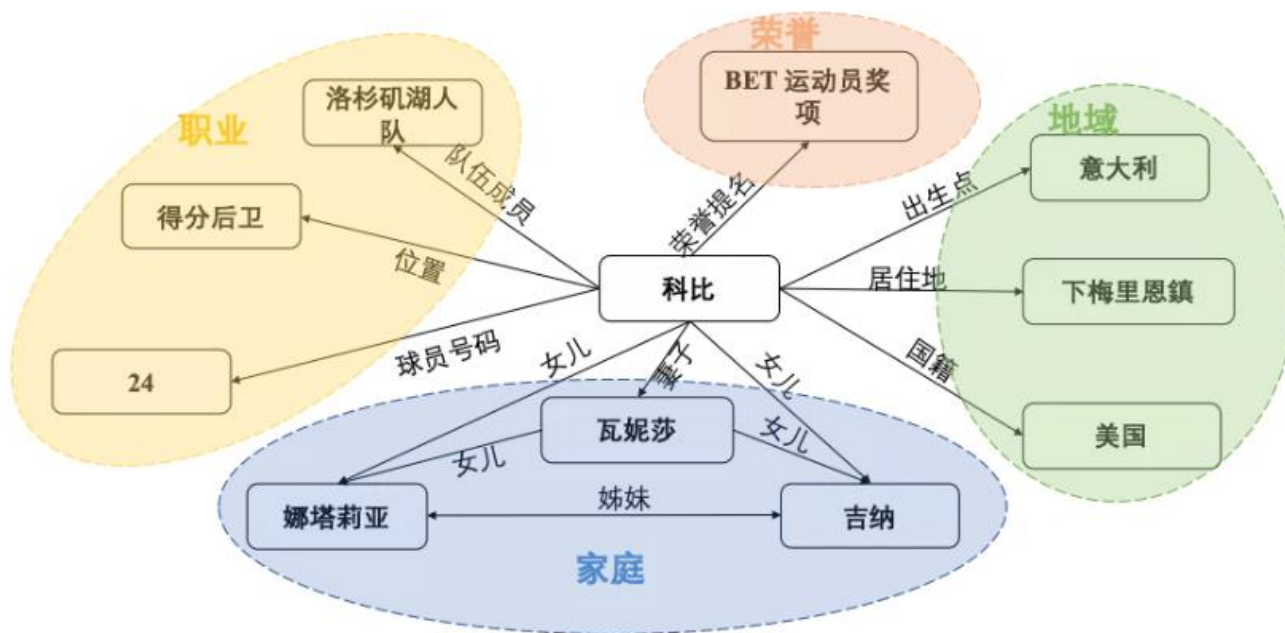
知识
图谱
补全

解耦
表征
学习

研究背景

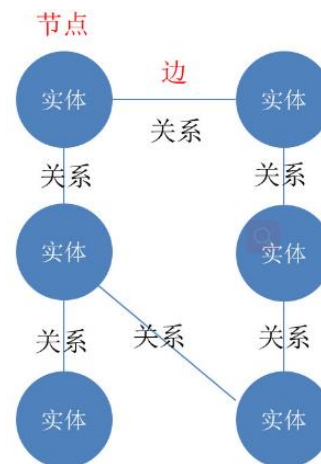


研究背景



知识图谱

- 知识图谱(KGs)被定义为一个有向图，它存储了真实世界实体和事实的结构化信息。
- 设 $G = (V, R, L)$ 为KG的一个实例，其中 V 、 R 和 L 分别表示实体(节点)集、关系集和边(事实)集。每条边 $e \in L$ 都呈现为一个三重 $(h, r, t) \in V \times R \times V$ ，描述了从头部实体 h 到尾部实体 t 之间的关系。



互信息

- 互信息是衡量两个随机变量相互依赖程度的度量。随机变量 X 和 Z 之间的互信息定义如下:

$$I(x, z) = \mathbb{E}_{(p(x, z))} \left[\log \frac{p(x, z)}{p(x)p(z)} \right]$$

其中 $p(x, z)$ 为 x 和 z 的联合概率分布， $p(x)$ 和 $p(z)$ 为相应的边际分布。互信息能够捕获变量之间的非线性依赖关系，因此可以作为真正依赖关系的标准度量。

问题定义

给定一张残缺知识图谱，期望通过对其进行学习以实现残缺边的预测，如给定头实体 h 和边 r 进行真实尾实体的预测 $(h, r, ?)$ 。值得注意的是，现有任务往往是转变为一个排名任务，即期望通过打分函数 $\psi(h, r, t) : \mathcal{V} \times \mathcal{R} \times \mathcal{V} \rightarrow \mathcal{R}$ ，对真实样本的预测情况优于负样本。

一些相关的工作将这个问题表述为 $(?, r, t)$ ，在给定一个尾实体和一个关系的情况下，推断缺失的头实体。事实上，这两个问题可以归结为同一个问题——学习一个合适的分数函数：

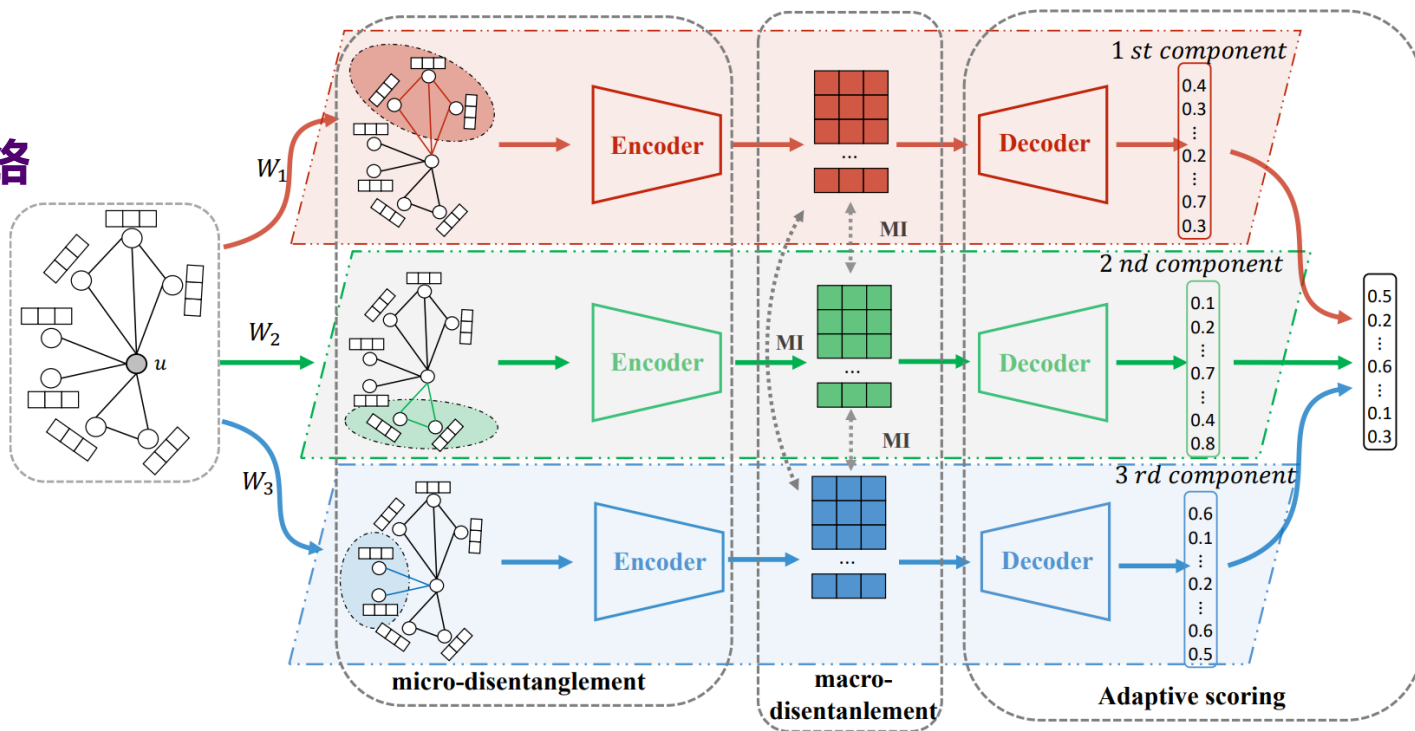
$$\psi(h, r, t)$$

基于解耦表征学习的图谱补全框架

01 关系感知网络

02 独立性约束

03 自适应评分



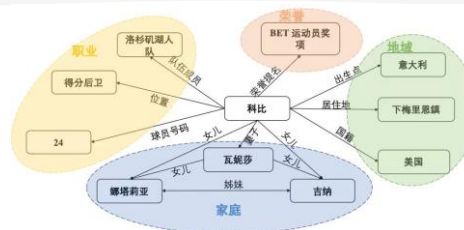
最终的结果需要根据各通道的预测结果再根据当前场景与各通道的相似程度判断给出最终的预测分数排名。

➤ 模块一 解耦转换

本研究提出的模型希望学习出的表征能够**实现解耦**，即每部分表征代表不同的含义。具体来说，对于实体 u ，假设它的表征由 K 个独立因子共同决定，如 $e_u = [h_{u,1}, h_{u,2}, \dots, h_{u,k}]$ ，其中 $h_{u,k} \in \mathbb{R}^{\frac{d_{\text{embed}}}{K}}$ ，表示实体 u 在第 k 部分的表征。为了实现上述目标，作者首先将初始的实体表征矩阵**映射到不同的隐空间**当中，从而更好地帮助挖掘节点特征中的不同语义。

$$h_{u,k}^0 = \sigma(W_k \cdot x_u)$$

其中，实体的初始化表征是通过各实体表征通过 K 个独立的映射矩阵 $W = \{W_1, W_2, \dots, W_K\}$ 得到的， σ 是非线性激活函数， x_u 是节点的特征。



➤ 模块二 微观解耦的关系感知聚合

- 在聚合邻居信息时仅使用与当前主题最为相关的部分，并非所有邻居实体。在聚合过程中应当显式融入邻居边信息以更好地实现解耦。

$$m_{(v,k,r)} = \phi(h_{v,k}, h_r, \theta_r)$$

$$\theta_r = W_r = \text{diag}(w_r)$$

- 为了更好地捕捉实体 u 与实体 v 之间关于子表征 k 的相关性，提出一个关系感知注意力机制来衡量两实体该部分重要性程度。

$$\begin{aligned} \alpha_{(u,v,r)}^k &= \text{softmax}((e_{u,r}^k)^T \cdot e_{v,r}^k) \\ &= \frac{\exp((e_{u,r}^k)^T \cdot e_{v,r}^k)}{\sum_{(v',r) \in \hat{N}(u)} \exp((e_{u,r}^k)^T \cdot e_{(v',r)}^k)} \end{aligned}$$

- 在得到上述注意力分数后，我们可以从邻居实体中聚合信息并且更新中心实体各子部分表征：
$$h_{u,k}^{l+1} = \sigma\left(\sum_{(v,r) \in \hat{N}(u)} \alpha_{(u,v,r)}^k \phi(h_{v,k}^l, h_r^l, \theta_r)\right)$$

➤ 模块三 宏观解耦的独立性约束

- 在复杂多关系图中，期望**不同语义空间中的子表征相互尽可能独立**，即降低彼此直接依赖。
- 本文采用**互信息最小化对不用语义间相关性进行控制**。具体采用一种对比对数比上界的MI估计器来实现解耦。核心思想为借助对比正负样本之间的差异从而对互信息上界进行估计。

$$\mathcal{L}_{mi} = \sum_i \sum_j \mathbb{E}_{(h_{u,i}, h_{u,j}) \sim p(h_{u,i}, h_{u,j})} [\log q(z_{u,i} | z_{u,j})] \\ - \mathbb{E}_{(h_{u,i}, h_{u',j}) \sim p(h_{u,i}) p(h_{u,j})} [\log q(z_{u,i} | z_{u',j})]$$

$$\mathcal{L}(h_{u,i}, h_{u,j}) = \mathbb{D}_{KL}[p(h_{u,i} | h_{u,j}) || q_{\theta}(h_{u,i} | h_{u,j})]$$

➤ 模块四 自适应评分

- **语义嵌入水平预测：** 打分函数本文采用现有三类经典打分函数。计算每个候选三元组 (u, r, v) 在每个组件中的分数。

$$\psi_{(u,r,v)}^k = f(\text{vec}(f(\overline{h_{u,k}^L}; \overline{h_r^L} \star \omega))W)h_{v,k}^L$$

- **关系感知融合：** 为了使本文模型能够适应于给定模型，在解码器部分之后加入了注意力评分模块 $\beta_{(u,r,v)}^k$ 。假设在关系感知语义子空间当中，最佳匹配当前语义环境的语义表征应当“更接近于”该语义空间中连接边的表征。

$$\begin{aligned}\beta_{(u,r)}^k &= \text{softmax}((h_{u,k}^L \circ \theta_r)^T \cdot h_r^L) \\ &= \frac{\exp((h_{u,k}^L \circ \theta_r)^T \cdot h_r^L)}{\sum_{k'} \exp((h_{u,k'} \circ \theta_r)^T \cdot h_r^L)}\end{aligned}$$

$$\psi_{(u,r,v)}^{\text{final}} = \sum_k \beta_{(u,r)}^k \psi_{(u,r,v)}^k$$

➤ 损失函数

- 在训练过程中，本文利用标准交叉熵损失与标签平滑，损失定义如下：

$$\mathcal{L} = -\frac{1}{B} \frac{1}{N} \sum_{(u,r) \in batch} \sum_i (t_i \cdot \log(\psi_{(u,r,v_i)}^{final}) \\ + (1 - t_i) \cdot \log(1 - \psi_{(u,r,v_i)}^{final})) + \lambda \cdot \mathcal{L}_{mi}$$

式中B为批量大小，N为知识图谱的实体数， t_i 为给定查询(u,r)的标签， \mathcal{L}_{mi} 和 λ 为互信息正则化损失及其对应的超参数。

01 数据集

常用公共数据集: FB15k-237、WN18RR

Data sets	$ \mathcal{E} $	$ \mathcal{R} $	Triplets		
			Train	Valid	Test
FB15k-237	14,541	237	272,114	17,535	20,466
WN18RR	40,943	11	86,835	3,034	3,134

02 对比模型

- 基于距离的模型 (TransE, RotatE)
- 语义匹配模型 (Distmult, RESCAL)
- 基于神经网络的模型 (ConvE、InteractE、SACN、ArcE、ReinceptionE、COMPGCN)

03 评价指标

平均倒数排名 (MRR)、平均排名 (MR)、一命中率 (Hits@1)、三命中率 (Hits@3)、十命中率 (Hits@10)

04 参数设置

组件嵌入大小: 200、优化器: Adam

Model	FB15k-237					WN18RR				
	MRR	MR	Hits@1	Hit@3	Hit@10	MRR	MR	Hits@1	Hit@3	Hit@10
TransE [5]	0.294	357	-	-	0.465	0.226	3384	-	-	0.501
Distmult [42]	0.241	254	0.155	0.263	0.419	0.43	5110	0.39	0.44	0.49
ConvE [8]	0.325	244	0.237	0.356	0.501	0.43	4187	0.40	0.44	0.52
RotatE [31]	0.338	<u>177</u>	0.241	0.375	0.533	0.476	3340	0.428	0.492	0.571
SACN [29]	0.35	-	0.261	0.39	0.54	0.47	-	0.43	0.48	<u>0.54</u>
InteractE [34]	0.354	172	0.263	-	0.535	0.463	5202	0.43	-	0.528
MuRE [2]	0.336	-	0.245	0.370	0.521	0.465	-	0.436	0.487	0.554
COMPGCN [35]	0.355	197	0.264	0.39	0.535	<u>0.479</u>	3533	0.443	<u>0.494</u>	0.546
AcrE [27]	<u>0.358</u>	-	<u>0.266</u>	<u>0.393</u>	<u>0.545</u>	0.459	-	0.422	0.473	0.532
ReInceptionE [41]	0.349	173	-	-	0.528	0.483	1894	-	-	0.582
DisenKGAT	0.368	179	0.275	0.407	0.553	0.486	1504	<u>0.441</u>	0.502	0.578

- FB15k-237数据集包含 237 种类型边，在 F15k-237 数据集上效果突出
- 有效解决复杂多语义知识图谱表示补全的问题

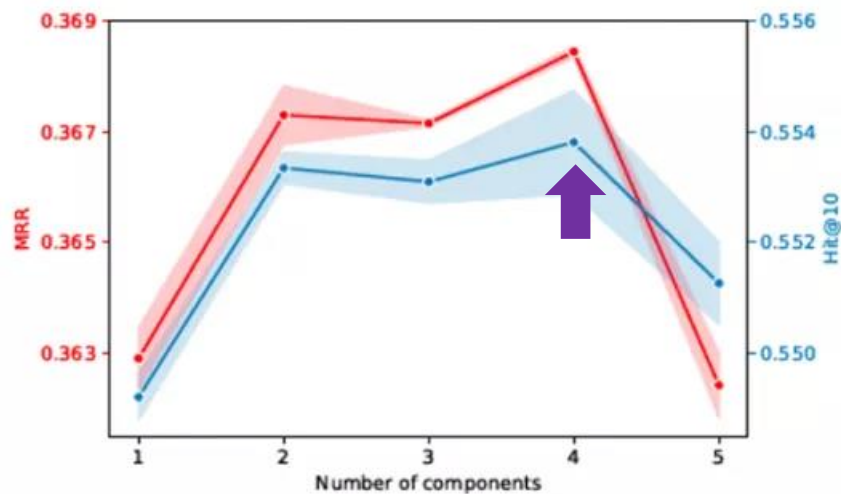
		RotatE		WGCN		COMPGCN		DisenKGAT	
		MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
Head Pred	1-1	0.498	0.593	0.422	0.547	0.457	0.604	0.501	0.625
	1-N	0.092	0.174	0.093	0.187	0.112	0.190	0.128	0.248
	N-1	0.471	0.674	0.454	0.647	0.471	0.656	0.486	0.659
	N-N	0.261	0.476	0.261	0.459	0.275	0.474	0.291	0.496
Tail Pred	1-1	0.484	0.578	0.406	0.531	0.453	0.589	0.499	0.641
	1-N	0.749	0.674	0.771	0.875	0.779	0.885	0.789	0.889
	N-1	0.074	0.138	0.068	0.139	0.076	0.151	0.086	0.180
	N-N	0.364	0.608	0.385	0.607	0.395	0.616	0.402	0.629

- 选取数据集：FB15k-237
- 四类关系：一对一 (1-1)、一对多 (1-N)、多对一 (N-1)、多对多 (N-N)
- 对于简单类型边、复杂类型边，模型效果均显著提升，大大优于其他模型

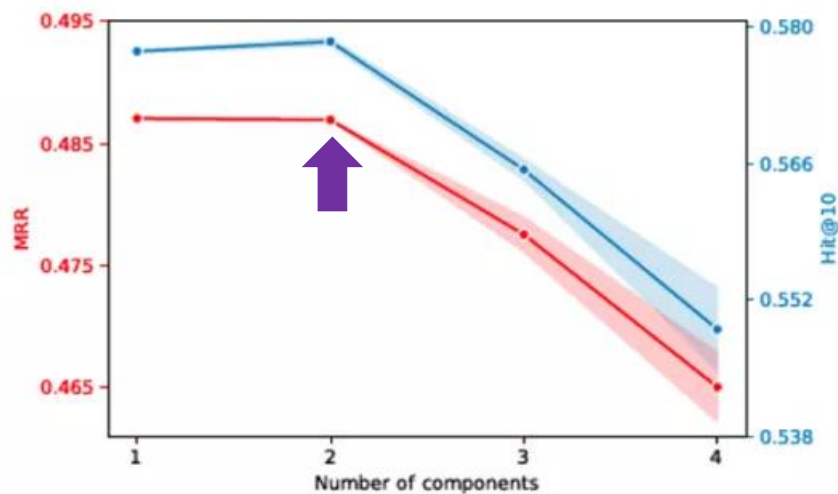
model	MRR	MR	Hits@1	Hits@3	Hits@10
w/o micro	0.355	197	0.265	0.392	0.534
w/o macro	0.356	303	0.263	0.392	0.542
w/o HSIC	0.352	259	0.263	0.387	0.527
DisenKGAT	0.368	179	0.275	0.407	0.553



- w/o micro 无微观解耦；w/o macro 无宏观解耦；w/o HSIC 更换其他独立性正则项
- 无宏观解耦，效果下降显著
- 更换其他独立性正则项，效果下降更为明显



(a) FB15k-237



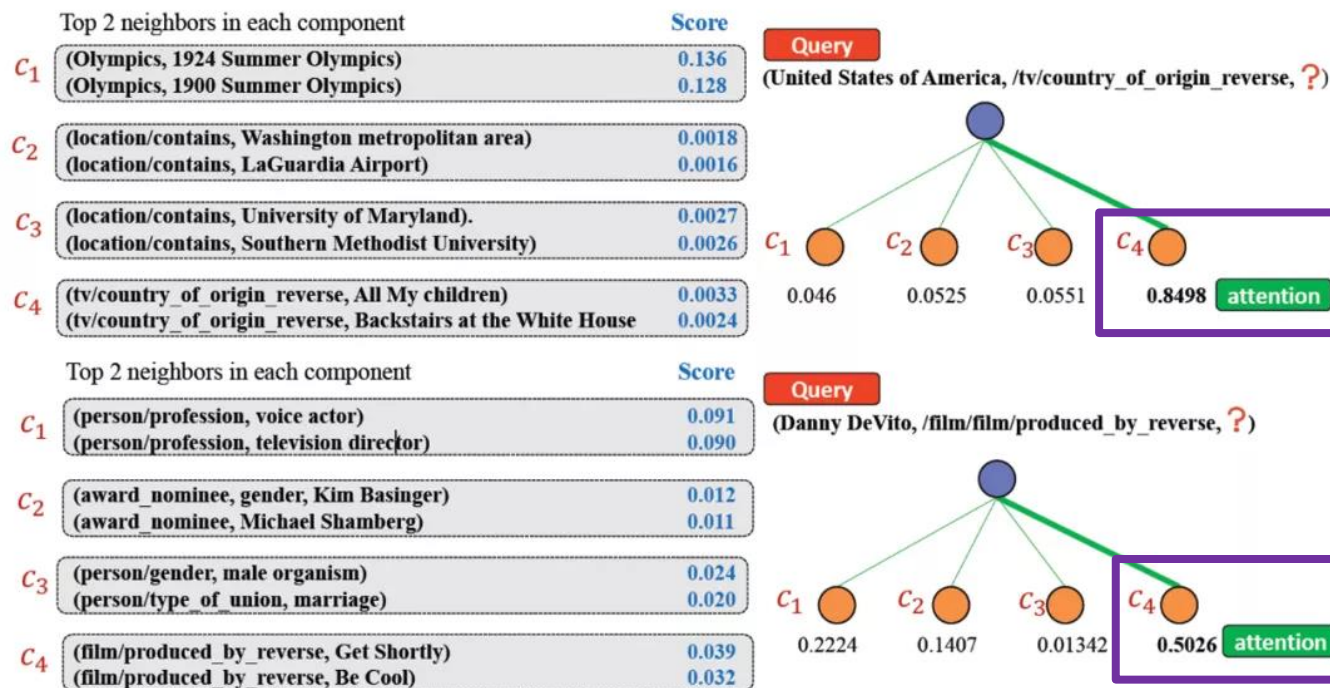
(b) WN18RR

- 对于FB15k-237 数据集，子语义 K 值呈现先升后降的趋势，最佳选择约为4
- 对于WN18RR 数据集，子语义 K 值增大效果持续下降，最佳选择约为 2

Scoring Function(=X)→	TransE			DistMult			ConvE		
Methods↓	MRR	MR	H@10	MRR	MR	H@10	MRR	MR	H@10
X	0.294	357	0.465	0.241	354	0.419	0.325	244	0.501
X+D-GCN	0.299	351	0.469	0.321	225	0.497	0.344	200	0.524
X+W-GCN	0.264	1520	0.444	0.324	229	0.504	0.244	201	0.525
X+COMPGCN(sub)	0.335	194	0.514	0.336	231	0.513	0.352	199	0.530
X+COMPGCN(Mult)	0.337	233	0.515	0.338	200	0.518	0.353	216	0.532
X+COMPGCN(Corr)	0.336	214	0.518	0.335	227	0.514	0.355	197	0.535
X+DisenKGAT(sub)	0.334	<u>183</u>	0.51	0.346	<u>196</u>	0.531	0.358	181	0.543
X+DisenKGAT(Mult)	<u>0.342</u>	170	<u>0.524</u>	<u>0.353</u>	184	<u>0.536</u>	<u>0.364</u>	171	<u>0.550</u>
X+DisenKGAT(Corr)	0.338	203	0.520	0.341	200	0.528	0.359	189	0.541
X+DisenKGAT(Cross)	0.343	187	0.526	0.354	204	0.540	0.368	<u>179</u>	0.553

- 定义四种算子： 减法 (Sub)、乘法 (Mult)、循环相关 (Corr)、交叉相互 (Cross)
- 具有很强的鲁棒性，可适配多种打分函数与融合方式，均有显著提升

DisenKGAT的可解释性



- 各个语义有较强的主题信息，不同主题具有显著差异性
- 由于不同语义信息彼此共享，两个相关联实体在部分语义主题一致



01 总结

- 知识图谱解耦表征学习框架 (DisenKGAT)
- 微观解耦与宏观解耦相结合

02 展望

- 探索更通用的解耦框架
- 在更多不同的场景探索互信息



谢谢聆听

Thank You