

基于拓扑结构与相似度信息融合的教育知识图谱节点重要性评估模型

李美子¹, 伍云芳¹, 卢淑怡¹, 王浩², 杨茹^{1,3}

1. 上海师范大学信息与机电工程学院, 上海 200234;

2. 上海新致软件股份有限公司, 上海 200127;

3. 上海师范大学上海智能教育大数据工程技术研究中心, 上海 200234

摘要

教育知识图谱是用于表示教育领域知识与概念之间关系的重要工具。在教育知识图谱中, 理解和评估节点的重要性对于教育资源管理、学习路径推荐等任务至关重要。然而, 传统的节点重要性评估方法通常将图中所有节点视为相同、只考虑某些单一拓扑结构, 不能充分挖掘知识图谱中节点的综合特征, 且忽视了实际中人们通常更关注重要性高的节点。为了解决该问题, 提出了基于重要节点驱动的拓扑结构与相似度信息融合的节点重要性评估模型 TSFM。该模型整合拓扑结构和由重要节点驱动的语义、结构上的相似度信息, 以评估教育知识图谱中节点的重要性。TSFM 利用图神经网络对知识图谱中的节点进行嵌入表示, 同时通过考虑节点之间的拓扑相似度和语义相似度来优化嵌入表示。实验结果表明, TSFM 在多个评价指标上表现出优异性能, 明显优于传统图算法和较新的图神经网络模型。

关键词

教育知识图谱; 节点重要性评估; 拓扑结构; 图注意力网络

中图分类号: TP302

文献标志码: A

doi:10.11959/j.issn.2096-0271.2025077

Node importance evaluation model for educational knowledge graph based on topological structure and similarity information fusion

Li Meizi¹, Wu Yunfang¹, Lu Shuyi¹, Wang Hao², Yang Ru^{1,3}

1. The College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China

2. Shanghai Newtouch Software Co., Ltd., Shanghai 200127, China

3. Shanghai Engineering Research Center of Intelligent Education and Bigdata, Shanghai Normal University, Shanghai 200234, China

Abstract

Educational knowledge graph is an important tool for representing relationships between knowledge and concept in the

field of education. Understanding and evaluating the importance of node in educational knowledge graph are crucial for tasks like educational resource management and learning path recommendation. However, traditional node importance evaluation methods often treat all nodes equally, considering only certain topological structures, and fail to capture the comprehensive characteristics of node. To address this, we proposed TSFM driven by key node. TSFM evaluated node importance by integrating topological structure with semantic and structural similarity information driven by key node. Specifically, TSFM utilized graph neural networks to embed nodes in knowledge graph, and optimized the embedding representation by considering topological and semantic similarity between nodes. Experimental results demonstrate that TSFM outperforms traditional algorithms and recent graph neural network models across multiple evaluation metrics.

Key words

educational knowledge graph, node importance evaluation, topological structure, graph attention network

0 引言

随着大语言模型（以下简称大模型）与知识图谱深度融合的发展趋势，知识图谱的构建与应用正逐步转向自动化、语义增强与多领域适配的新阶段。大模型不仅提升了知识图谱的生成效率，还借助知识图谱增强自身的语义推理能力，实现了“知识增强的大模型”与“模型驱动的知识图谱”的双向赋能^[1]。在此背景下，教育领域亟须融合结构化图谱与深度学习模型，提升知识点重要性评估的科学性与智能化水平。随着现代教育领域技术的进步和教育需求的不断演变，知识图谱已成为连接散落知识点、提高教学与学习效率的关键工具^[2]。教育知识图谱作为知识图谱在教育领域的具体应用，旨在构建基于教育内容和知识点的语义网络，支撑个性化学习与智能教学等多种教育场景^[3]。已有研究围绕其概念模型、构建方法与应用展开了深入探讨，普遍认为教育知识图谱不仅可以表征学科知识点之间的逻辑与语义关系，还可以服务于认知诊断、路径推荐与资源整合^[4-6]。尽管学术界尚未对其形成统一定义，但主流观点主要集中于两类：一是整

合学科知识与学习者认知状态，支持个性化学习分析；二是聚焦于结构化知识点的层次建模，强调知识点间的依赖关系和拓扑结构^[7]。本文采用第二类观点，以教育知识点的结构关联建模为核心，结合拓扑结构与语义信息进行节点重要性评估，不涉及个体学习状态等外部特征。

教育知识图谱通过整合多学科的知识结构^[8]，为教师和学生提供一个全面、相互关联的知识框架，这一过程标志着教育技术从传统的“感知智能”向更高级的“认知智能”转变。随着教育技术从“感知智能”向“认知智能”的转变，教育知识图谱在教育数字化转型中日益发挥关键作用。为进一步推动其发展，全球研究人员正从知识建模、路径导航、认知学习和资源管理等多个维度积极探索其应用价值。

知识图谱中的“节点”通常表示知识点或概念，其在知识图谱中通过各种关系互相连接，如图1所示。其中，“函数”是一个数学概念节点，其通过“包含”边与“反函数”次级节点相连，通过“部分元素”边与“定义域”“对应关系”“值域”次级节点相连。每个节点还可能与相关的数学定理、公式或知识点的应用相连接。例如，“单调性”节点下有易混淆的应用知识点“在某个区间上单调递增”和“某个

传统方法包括度中心性、接近中心性和介数中心性等。例如, Qiong等^[9]提出了一种基于节点偏心率的评估方法; Xu^[10]提出了一种结合多种中心性指标的因子分析改进TOPSIS方法; Ai等^[11]提出基于网络熵变化的方法, 通过节点移除前后的结构扰动衡量其全局重要性; Le等^[12]基于数据场理论, 引入拓扑势概念以识别关键节点。考虑到节点本身及其邻居节点的度对节点重要性的影响, Hu等^[13]引入节点间存在相互依赖关系的观点。PageRank^[14]是该观点的经典算法, 最初是为了评估网站的重要性而设计, 但其仅基于图结构, 不能评估某些节点的重要性分数。Haveliwala^[15]在此基础上提出PPR (personalized PageRank), 通过允许用户在图中提供自己的节点重要性突破了这一局限, 但PPR不能考虑边缘类型。RWR (random walk with restart)^[16]进一步增强了对起始节点局部结构的捕捉能力, 但依然难以应对异构图与语义复杂图的挑战。HAR (hub, authority and relevance)^[17]以区分知识图谱中不同谓词, 同时注意重要性分数和图拓扑结构来拓展PageRank和PPR。上述方法侧重于图的拓扑结构, 并普遍假设越重要的节点通常会获得更多其他节点的连接, 从而依据连接数量对节点的重要性进行初步估算, 因此该方法在复杂图结构下的表达能力存在一定局限。同时上述方法忽视了图中编码的实质性语义和潜在结构信息, 在应用于大规模复杂知识图谱中时仍存在不足。

随着技术的发展, 基于图结构的数据(如社交网络、推荐系统和知识图谱等)逐渐应用于多个领域。为了更有效地处理和分析复杂的图结构数据, 图学习方法(如图神经网络^[18])应运而生, 逐渐成为图结构数据分析与建模领域中的关键技术工具。

图学习方法不仅提升了对图结构和特征的理解, 还为节点重要性评估提供了更加灵活、准确的方法。在此背景下, 近年来涌现了多种基于图神经网络的可训练方法, 如GENI (graph neural network for estimating node importance)^[19]、Multiimport^[20]和RGTN (relational graph transformer network)^[21], 以进一步增强节点重要性评估效果。GENI基于node2vec^[22], 通过捕捉局部邻居信息生成节点嵌入表示, 并引入图注意力网络(graph attention network, GAT)^[23]聚合不同类型的边信息, 提升了模型对复杂节点关系的捕捉能力。Multiimport在此基础上结合外部输入的重要性信号, 提高了模型的灵活性, 使其能够适应不同的应用场景, 尤其在跨学科场景下表现优异。RGTN通过双向编码器同时捕获节点的结构和语义信息, 在大规模复杂图结构下表现出色。HIVEN^[24]进一步引入元模式以缓解节点类型偏倚, 增强了模型在异构图中处理节点相似性的能力, 但在超大规模图上的可扩展性仍有限。随着图结构数据在教育等领域的广泛应用, 图神经网络逐渐成为节点重要性建模的重要工具。代表性方法如图神经网络通过邻居特征聚合构建节点表示, 但难以处理大规模图与节点差异性。GraphSAGE^[25]引入邻居采样机制提升效率, 仍未有效区分邻接节点的重要性。GAT^[23]借助注意力机制实现了动态加权聚合, 在异构图中开销较大。GIN^[26]增强了结构判别能力, 但扩展性仍有限。为适应异构图建模, HetGNN^[27]与RGCN^[28]引入节点类型与关系建模, 显著提升了语义区分力。然而, 上述方法多聚焦于结构建模, 对节点语义相似性与动态图谱中的演化机制融合不足, 难以全面反映教育知识图谱中的跨学科、多层次关联特征。

尽管已有的知识图谱节点重要性评估方法取得了一定成果，但教育知识图谱是专门用于描述教育领域中知识点及其相互关系的复杂网络，具有层次化结构、丰富的语义关联和动态演化性等特点，因此适用于普通知识图谱的方法并不能较好地直接应用于教育知识图谱中。具体来说，首先，基于统计指标的传统方法主要基于节点的拓扑结构（如度数、路径长度）进行评估，然而在教育知识图谱中，语义关联往往比结构连接更能反映知识点的重要性。由于未能有效整合语义信息，现有方法在评估精度上存在明显不足。其次，教育知识图谱中的多层次结构和不同学科的复杂交互，使节点重要性评估更加多样化和复杂化。然而，现有方法往往对所有节点采用相同的评估标准，忽视了不同层次或不同学科节点的重要性差异，无法针对性地评估不同教育阶段的知识需求。为了解决上述问题，本文提出了一种新的教育知识图谱节点重要性评估模型 TSFM（topological structure and similarity information fusion model）。首先，针对节点评估中过度依赖拓扑结构的问题，本文综合考虑了节点的语义信息，使 TSFM 评估结果不仅基于节点间的结构关联，还能够反映节点之间的语义关联。其次，TSFM 能够满足不同学科和不同教育层次的多样化需求，通过由重要节点驱动的评估标准，确保在跨学段、多学科的知识图谱中实现更加精确的节点重要性评估。本文主要贡献如下。

- 本文提出了一种基于图神经网络的节点重要性评估模型 TSFM。该模型不仅结合了拓扑结构，还通过融合节点的语义相似度，进一步优化了节点嵌入表示，能够更准确地评估图中节点的重要性。

- 本文聚焦于教育知识图谱中的重要节点，通过比较重要节点与其他节点在结

构和语义信息方面的相似度，结合结构信息与语义信息的相似度计算，动态调整节点评分，能够更准确地捕捉与重要节点具有高相似度的其他节点，进而提高节点重要性评估的精度。

- 本文在跨学段、多学科融合的数据集上对所提模型进行了实验验证，实验结果表明，本文提出的 TSFM 在多个评价指标上均优于现有方法。从而证明了其教育知识图谱中评估节点重要性时的有效性和实用性。

1 相关概念说明

知识图谱中节点重要性评估涉及的相关概念和符号如下。

（1）知识图谱

知识图谱是一种用可视化的手段来描述知识发展进程和知识结构关系的领域映射地图。将知识图谱形式化为有向图 $KG = (E, R, F)$ 。其中， $E = (e_0, e_1, \dots, e_n)$ 表示节点集合； $R = (r_0, r_1, \dots, r_n)$ 表示关系集合； $F = \{(h, r, t)\} \subseteq E \times R \times E$ 是关系三元组的集合。其中， h 表示头节点； t 表示尾节点； r 表示节点之间的关系，均为有向连接。

（2）教育知识图谱

教育知识图谱是知识图谱在教育领域当中的实际应用，本文基于知识图谱的概念定义教育知识图谱。将教育知识图谱形式化为有向图 $EKG = (E_{edu}, R_{edu}, F_{edu})$ 。节点集合 $E_{edu} = (c_0, c_1, \dots, c_n)$ 中的节点表示教育教学知识点，关系集合 $R_{edu} = (\text{include}, \text{precedes}, \dots, \text{applies})$ 中的关系包括包含、前驱、等价、推导、应用、相似等知识之间的连接关系。

（3）Top-K 排序

在教育知识图谱中，给定节点集合

$E = (e_0, e_1, \dots, e_n)$, 节点重要性评估模型 $f(\cdot)$ 通过计算节点的重要性并排序, 得到该教育知识图谱中重要性最高的 k 个节点。

(4) 中心度因子

在知识图谱中, 中介中心度指一个节点作为其他两个节点最短路径间桥梁的次数。一个节点的中介中心度越大, 表示该节点在不同节点之间充当纽带的次数越多。在学科知识系统中, 重要的知识点在知识结构中能够充当纽带, 且有助于解决具体的教学内容和教学目标, 并提供直接支持。因此, 本文将节点的中介中心度作为中心度因子, 通过计算中心性因子, 可以准确描述知识点在整个知识图谱中发挥的纽带作用。

中心度因子 C_i 的具体计算式如式 (1) 所示。

$$C_i = \sum_{j,k \neq i} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \quad (1)$$

其中, σ_{jk} 是节点 j 和节点 k 之间的最短路径数, $\sigma_{jk}(i)$ 是经过节点 i 的最短路径数。最短路径通过 Dijkstra 算法^[29] 计算。

(5) 出入度因子

出入度因子指一个节点在知识图谱中出度和入度的总和。出度指从该节点指向其他节点的边的数量, 入度指其他节点指向该节点的边的数量。出入度因子越高的节点表示在知识图谱中有越多的输入和输出关联, 说明该节点在知识网络中具有越广泛的联系和影响力。在学科知识系统中, 出入度因子能够反映一个知识点的综合活跃程度和连接密度, 有助于评估知识点在知识图谱中的重要性和信息传播能力。因此, 出入度因子是衡量知识图谱中节点重要性的重要指标之一。出入度因子的具体计算式如式 (2) 所示。

$$\begin{cases} d'_i = w \times d_i \\ d'_o = w \times d_o \end{cases} \quad (2)$$

其中, d_i 表示入度; d_o 表示出度; w 表示连接的权重, 用于调整节点的入度 d_i 和出度 d_o 对最终评分的影响。

出入度因子反映了知识点在教育知识图谱中的相对重要性。较高的 w 赋予连接较大的权重, 增强高连接度节点的重要性; 较低的 w 则削弱度的影响, 使模型更加关注其他信息来源 (如语义信息)。需要注意的是, 本文实验部分通过网格搜索法对 w 进行调优, 最终选择 $w = 0.5$, 以确保拓扑结构信息与其他特征 (如语义相似度) 之间的平衡。其使用度信息既能发挥作用, 又不会在最终评分中占据主导地位, 从而提高模型对节点重要性的评估准确性。

(6) 通过率因子

通过率因子用于衡量一个知识点在随机游走过程中被遍历的次数, 以反映其教育知识图谱中的重要性。重要的知识点通常在长期学习体系中具有更深远的教育意义。因此, 评估知识点的遍历次数可以表明其对学生学习的影响。为了更加准确地计算通过率因子, 本文结合随机游走和最短路径算法来分析知识点的转移概率和关联性。

在教育知识图谱 $EKG = (E_{edu}, R_{edu}, F_{edu})$ 中, 定义一个长度为 $l \in N$ 的节点遍历序列 $(v_0, v_1, \dots, v_l) \in V^{l+1}$ 。其中, 满足 $v_i \in E_{edu}$, $\forall i \in |l|$ 。随机游走从某个随机选择的知识节点 $v_0 \in V$ 开始, 以均匀分布策略从邻居节点集合 $N_{KG}(v_i)$ 中随机采样下一个节点 v_{i+1} , 最终得到节点间的转移概率矩阵 P , 该矩阵反映了节点间的可达性和关联性。

通过率因子 T_i 的计算式如式 (3) 所示。

$$T_i = L_i * p_i \quad (3)$$

其中, L_i 表示临界度, p_i 表示节点的转移

概率。

临界度 L_i 用来衡量节点在连接路径中的控制能力。

$$L_i = \begin{cases} 0, & d=0 \\ \frac{\sum_{j,r \in N_{kg}} \frac{1}{q_{jr}}}{\sum_{j,r \in N_{kg}} \frac{1}{q_{jr}} + \sum_{j,r \in N_{kg}} \frac{1}{k_{jr}}}, & d>1 \end{cases} \quad (4)$$

其中, d 表示节点的度; q_{jr} 表示在不经过节点 i 的情况下, 从节点 j 到节点 r 的最短路径条数; k_{jr} 表示在经过节点 i 的情况下, 从节点 j 到节点 r 的最短路径条数。

(7) 节点重要性评估

给定一个教育知识图谱 $EKG = (E_{edu}, R_{edu}, F_{edu})$, 和重要节点子集 $V_s \in V$ 的部分已知节点重要性辅助得分 $\{I\}$ 。节点重要性估计旨在学习函数 $F: E_{edu} \rightarrow [0, \infty)$, 估计教育知识图谱中每一个节点的重要性评

分 $s \in \mathbb{R} (\mathbb{R} \geq 0)$, 并对 s 进行排序。核心问题可形式化建模为 $F: KG \Rightarrow s$ 。

数据集中存在缺乏完整重要性评分信息的节点时, 需要手动或基于某些启发式规则选取一个重要节点子集 V_s , 用于指导模型的训练或评估。而本文使用的数据集中, 每个节点均包含初始重要性评分, 因此模型在评估时不依赖于特定的预选子集 V_s , 可以直接基于所有节点的结构与语义信息进行全局重要性计算。

2 模型介绍

2.1 模型整体架构

图2所示为本文提出的节点重要性评估模型 TSFM 的整体架构, 包括3个主要

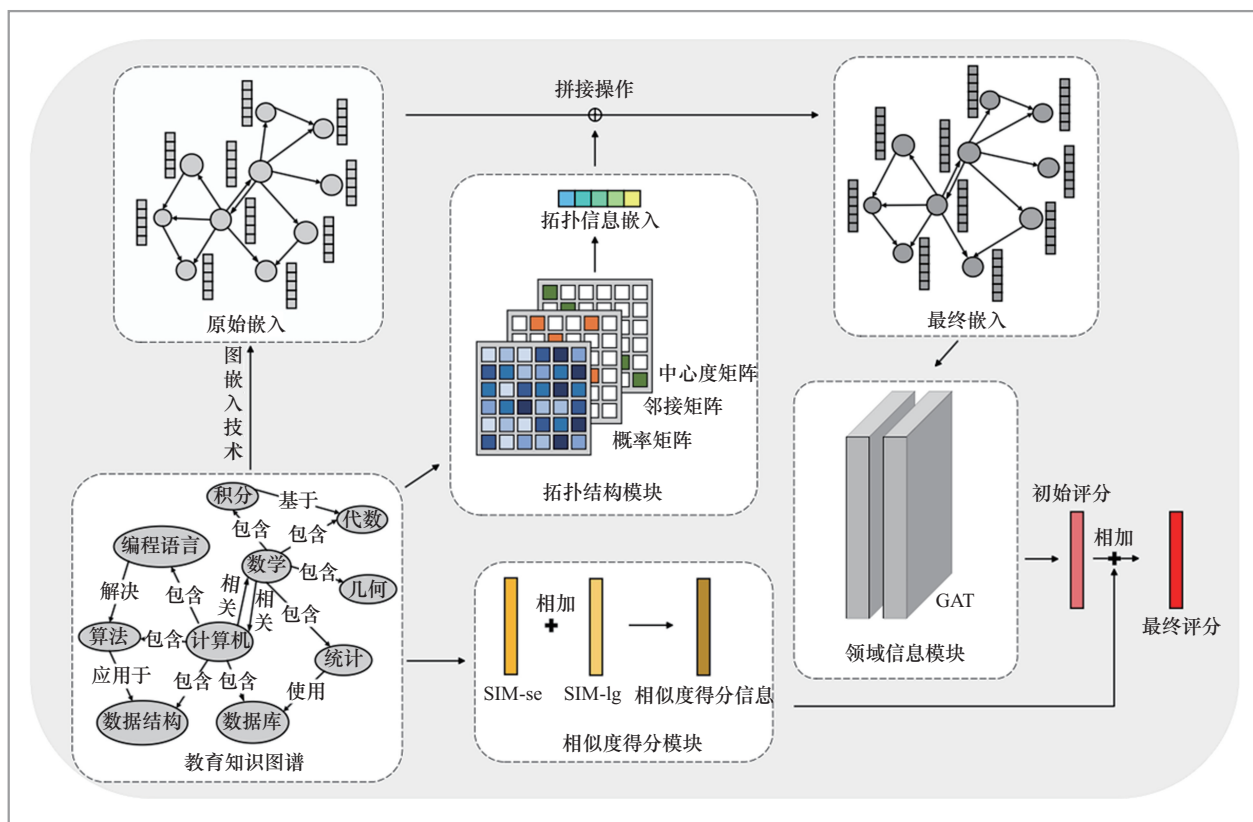


图2 TSFM的整体架构

部分：拓扑结构模块、相似度得分模块以及邻域信息聚合模块。该模型通过整合拓扑结构信息、相似度得分信息和邻域信息，增强其在教育知识图谱中识别关键节点的能力。

首先，模型从教育知识图谱中获取原始嵌入。教育知识图谱包括各种计算机科学和数学的节点，如编程语言、算法、数据结构、数据库、数学、统计学等，这些节点通过不同的关系，如“包含”“基于”“解决”相互连接。基于这些连接，模型使用图嵌入技术生成节点嵌入，并为知识图谱中的节点提供高维向量表示。其中，图嵌入技术^[30]是一种基于复数空间的知识图谱嵌入方法，通过旋转变换来建模节点之间的关系。在图嵌入技术计算框架下，模型将每一个三元组 (h, t, r) 中的每个关系 r 建模为复数空间中的旋转变换，即 $t = h \circ r$ ，其中， $\circ$ 表示复数空间中的逐元素乘法，用于模拟头实体向尾实体的旋转映射。相比于传统的 TransE^[31]和 DistMult^[32]，图嵌入技术更擅长处理具有复杂关系结构的知识图谱，因此适用于教育知识图谱中的节点表示学习。

然后，在原始嵌入基础上，拓扑结构模块使用图的拓扑信息来进一步优化节点嵌入。该模块利用中心度矩阵（反映节点的全局影响力）、邻接矩阵（表示节点之间的直接连接关系）以及基于路径概率构建的转移矩阵，将上述拓扑结构信息整合进节点嵌入表示中。通过整合，每个节点的嵌入能够准确反映其在图中的结构重要性和连接模式。同时，相似度评分模块评估图中节点与重要节点之间在语义和结构信息方面的相似性，从而进一步细化节点评估得分。相似度评分模块采用两种不同的相似度信息 SIM-se 和 SIM-lg，SIM-se 侧重于评估节点间语义相似性，SIM-lg 侧重于计算节点间局部图结构相似性以加强节点得分评估的区分度。在邻域信息聚合

模块中，采用图注意力网络 GAT 来聚合节点的邻域信息。该模块考虑到每个节点的邻居特性，通过加权邻居的嵌入向量来更新节点的嵌入，聚合过程中使用的注意力机制确保了模型能够重视与当前节点相关性高的邻居节点，从而得到初始的节点重要性得分向量。

最后，将相似度评估模块得到的各节点相似度评分与初始节点重要性得分向量相加得到最终的预测结果，最终的节点重要性评分表示各节点在教育知识图谱中的重要性。

2.2 拓扑结构量化计算

在教育知识图谱的节点重要性量化计算中，本文利用教育知识图谱节点的中心度因子、出入度因子、通过率因子来优化节点嵌入。这些指标综合反映了节点在教育知识图谱中的角色和作用。

首先，中心度因子通过统计节点在所有最短路径中的出现频率确定每个节点的中心性得分，揭示其在知识网络中的中介作用。然后，模型通过计算出入度因子来考察节点与其他节点的连接程度。对于有向图，分别计算出度和入度，并引入权重来调整不同连接类型的影响力。中心度因子揭示了节点在知识体系中的基础性和枢纽性，以识别对学科理解具有重要作用的知识点。最后，模型引入通过率因子来衡量节点在教育知识图谱中的信息传递能力。该因子结合了节点的转移概率与临界度指标，综合反映其在图中所处位置与功能作用，从而对节点在教育知识图谱中的作用与重要性做出更全面的评估。为了进一步获得更准确的嵌入表示，模型在分别计算出中心度、出入度和通过率3类拓扑信息矩阵后，将这些结构性特征与初始节点嵌

入表示进行拼接操作，得到最终的节点嵌入表示。拓扑结构量化计算的整体过程如**算法1**所示。

算法1 拓扑结构量化计算

输入：图 $G=(V,E)$ ，其中 V 为节点集合， E 为边集合

输出：节点嵌入矩阵

1 初始化：初始化每个节点 $v \in V$ 的节点嵌入矩阵

2 函数：计算中心度因子

3 对于每个节点 $v \in V$ ，基于最短路径计算中心度因子 C_i

4 返回中心度因子 C_i

5 函数：计算出入度因子

6 对于每个节点 $v \in V$ ，分别计算出度因子 d'_i 和 d'_o ，并根据边的不同类型加权计算

7 返回出入度因子 d'_i 和 d'_o

8 函数：计算通过率因子

9 计算转移概率矩阵 P 和临界度 L_i ，然后计算通过率因子 $T_i = L_i \times p_i$

10 返回通过率因子 T_i

11 将每个节点的中心度因子、出入度因子和通过率因子与初始节点嵌入 H_v 进行拼接

12 输出：最终节点嵌入矩阵 H'_v

2.3 领域信息聚合

在许多场景中，节点的重要性不仅取决于其自身的特征，还受到邻居节点信息的显著影响。其在教育知识图谱中非常重要，因为知识点之间的关系和依赖性直接影响学习路径的设计和知识点的优先级。

若邻居节点在知识网络中扮演着关键角色，该节点一般在相关学习路径或知识关联中也具有较高的权重。GAT 在处理图相关的任务中展示了其强大的性能，因此本文利用其对节点的动态邻域进行信息聚合。

首先，本文利用一个可学习的函数 $g(\cdot)$ ，将每个节点的特征向量 x_i 投影到低维空间中，得到 $x'_i = g(x_i)$ （即将节点嵌入对齐到一个统一的空间，为后续操作提供便利）；进一步，将节点的表示细化为一维嵌入，通过函数 $g'(\cdot)$ 直接反映节点的重要性，表示为 $h_i^0 = g'(x'_i)$ ，其中 $g(\cdot)$ 、 $g'(\cdot)$ 为简单的线性变换函数。

然后，通过转换的向量，应用注意力机制聚合邻域信息，以此权衡教育知识图谱中邻居节点的重要性，从而为每个节点 i 计算出新的重要性表示。聚合嵌入 h_i^l 作为邻居嵌入的加权和计算，其中权重由注意力机制决定，反映了不同邻居的不同重要性。具体的计算公式如式（5）所示。

$$h_i^l = \sum_{j,m \in N_{KG}} \omega_{(i,j,m)}^l h_j^{l-1} \quad (5)$$

其中， ω 表示注意力机制中针对节点 i 与邻居节点 j 之间的局部权重系数。

考虑到教育知识图谱是一个具有多种节点间关系类型的有向异构图，本文需要基于关系类型而非节点本身计算注意力权重。在第 l 层聚合时，对于节点 i 与其邻居节点 j 之间的关系 m ，模型计算注意力权重 $\omega_{(i,j,m)}^l$ 。该设计使模型能够根据不同关系类型，在聚合过程中动态调整各邻居的贡献程度。 $\omega_{(i,j,m)}^l$ 的权重计算式如式（6）所示。

$$\omega_{(i,j,m)}^l = \frac{\exp \left(\text{LeakyReLU} \left(a_i^T \left[h_i^{l-1} \parallel \pi \left(\rho_{ij}^m \parallel h_i^{l-1} \right) \right] \right) \right)}{\sum_{(k,n) \in N_{KG}} \exp \left(\text{LeakyReLU} \left(a_i^T \left[h_i^{l-1} \parallel \pi \left(\rho_{ik}^n \parallel h_k^{l-1} \right) \right] \right) \right)} \quad (6)$$

其中, ρ_{ij}^m 表示节点 i 和节点 j 的关系类型; LeakyReLU 是带泄露的线性整流函数; π 是函数符号, 用来表示对拼接后的关系特征和节点特征进行映射或投影。

最后, 通过聚合嵌入应用非线性变换 (如 Relu) 得到节点重要性的初始表示 s_o , 用于综合反映节点在图结构中的整体重要性。

$$s_o = \text{ReLU}(h_i^l) \quad (7)$$

2.4 相似度得分计算

为了更精确地评估节点的重要性, 本文基于初始嵌入矩阵和拓扑结构量化矩阵计算相似度得分, 根据所有节点的重要性分数进行排序, 选取排名前 50 的节点作为高重要性节点, 并选取知识图谱中最重要的节点作为基准。

相似度得分模块引入了基于节点上下文的节点相似度函数, 旨在进一步得到图中各节点与图中所有重要节点在语义与结构方面的相似性。具体地, 引入的相似度函数包括两个主要部分: 节点语义相似度和节点的局部图相似度。对于节点 i 的相似度综合计算式如式 (8) 所示。

$$\text{SIM}_i = \frac{\text{SIM}_{\text{sc}}(i, j) + \text{SIM}_{\text{lg}}(i, j)}{|\mathcal{T}|} \quad (8)$$

其中, SIM_i 表示节点 i 和其他预先知道的重要节点的总体相似度; $|\mathcal{T}|$ 表示前 $\text{top}@k$ 个重要节点集合中的元素个数; $\text{SIM}_{\text{sc}}(i, t)$ 表示基于余弦相似度计算得到的节点语义相似度, 用于衡量节点间的语义关联程度, $\text{SIM}_{\text{lg}}(i, t)$ 表示节点的局部图相似度。

$$\text{SIM}_{\text{sc}}(i, t) = \sum_{t \in \mathcal{T}} \cos(i, t) = \frac{i \cdot t}{\|i\| \|t\|} \quad (9)$$

其中, $\cos(i, t)$ 量化了节点 i 和预先确认的前 $\text{top}@50$ 个重要节点集合 $\mathcal{T}$ 中各节点 t 间

的夹角余弦, 评估它们的语义接近程度。

节点的局部图相似度 $\text{SIM}_{\text{lg}}(i, t)$ 考虑了节点 i 和预先确认的前 $\text{top}@50$ 个重要节点集合 $\mathcal{T}$ 中各节点 t 之间在图中位置和关系属性之间的相似性。

$$\text{SIM}_{\text{lg}}(i, t) = \sum_{t \in \mathcal{T}} \frac{\text{val}(i) \cdot \text{val}(t)}{|\text{val}(i)| |\text{val}(t)|} \quad (10)$$

其中, $\text{val}(i)$ 和 $\text{val}(t)$ 分别表示节点 i 和节点 t 在局部拓扑结构中的嵌入向量, 由拓扑结构量化模块中生成的局部结构嵌入矩阵提取得到。该嵌入向量综合考虑了节点的中心度、出入度、通过率等结构特征, 用于刻画节点在图结构中的相对位置和连接属性。

经过上述计算得到的结构与语义相似度得分, 将与领域信息聚合模块生成的初始节点重要性得分进行融合, 最终输出节点的重要性评估结果。值得注意的是, 相似度得分模块的作用并非直接决定节点的重要性, 而是对初始重要性评分进行加权调整, 使最终评分更加符合知识网络的整体结构。

通过整合以上内容, 本文从以下两个角度进一步优化, 从而提高节点重要性评估得分的准确性。

(1) 优先考虑与内部重要性节点相似性高的节点

通过相似度函数 (包括语义相似度和局部图相似度) 优先选择与已确认的重要节点在语义和结构上具有高相似度的节点, 因为这些节点更有可能对知识图谱中的学习任务产生积极的影响。

(2) 最小化选择低相似度的节点

在筛选过程中, 避免选择与内部重要性节点相似度较低的节点, 以降低评估噪声; 有效过滤与重要节点不具有一致性的节点, 确保评估的精准性。

2.5 损失函数

在原始嵌入、拓扑结构模块与最终嵌入的基础上,模型通过优化目标函数来进行训练,以确保对节点重要性进行准确评估。为了进一步提高模型的评估节点重要性的性能,本文设计了一个针对节点特征的损失函数,旨在最大化重要节点与其他节点的区分度。本文所提出模型的损失函数定义如下。

$$\mathcal{L} = -\left(\sum_{i \in V} s'_i \log(s_i^*) + \alpha \sum_{i \in V} s'_i \log(\text{SIM}_i^{\text{sc}} \times s_i^*) + \beta \sum_{i \in V} s'_i \log(\text{SIM}_i^{\text{lg}} \times s_i^*)\right) \quad (11)$$

其中, s'_i 表示节点 i 的重要性真实值、 s_i^* 表示使用模型后预测的节点 i 的重要性值; $\sum_{i \in V} s'_i \log(s_i^*)$ 表示交叉熵损失函数; SIM_i^{sc} 、 SIM_i^{lg} 分别表示节点 i 与重要性节点集合 T 中各节点语义和结构相似性的加权和; α 、 β 均为超参数; $\text{SIM}_i^{\text{sc}} \times s_i^*$ 表示节点 i 在语义方面的得分情况; $\text{SIM}_i^{\text{lg}} \times s_i^*$ 表示节点 i 在结构方面的得分情况。

该损失函数引入两个加权项,分别针对语义和结构维度上的重要节点,通过抑制与重要节点差异较大的节点嵌入,增强模型在面对语义无关或结构不相似节点时的鲁棒性。

3 实验

3.1 数据集

目前,教育知识图谱很少有可用的公

开数据集。因此,本文收集和构建了一套教育知识图谱数据集,其包含了从初级教育到高等教育各个层次的数学学科知识点,以及大学计算机课程的学科知识点。每个知识点以节点形式存在,节点间的关系表示知识点之间的逻辑连接和依赖关系。为了确保数据集包含足够的拓扑复杂性和领域特异性,以模拟实际的教育环境,本文收集不同教育层次(从初级教育到高等教育)中数学学科的知识点和大学计算机课程的知识点,确保知识点之间的关系能够反映知识图谱中的多层次依赖结构;并通过仔细选择知识点之间的逻辑连接,确保它们能够反映实际教学中常见的关联方式和依赖链,形成与实际教育场景一致的图谱结构。数据集的简要描述见表1。

为了更好地让模型学习到重要节点,本文将多张教育知识图谱进行融合,找到不同教育知识图谱相连接的节点对。本文从粗粒度角度出发,利用知识图谱嵌入模型得到节点的语义信息嵌入;从细粒度角度出发,分析教育教学具有主题的特点,利用知识图谱嵌入方法(如 TransE、RotatE^[33]等)对节点的语义特征进行建模,生成对应的语义嵌入表示,以更好地表达知识节点的语义关联性、结构层次性和逻辑推理性。本文通过将语义嵌入与结构嵌入结合,综合其各自优势,通过多层感知机^[34]对融合嵌入向量进行降维处理,并计算节点间相似度,将相似度高的节点对进行对齐,从而构建一张大型跨学科教育知识图谱。其涵盖3 705个节点和

表1 数据集简要描述

数据集	节点数量/个	边数量/条	谓词数量/个	三元组数量/个
高中数学课程	2 399	1 1250	6	11 250
大学数学课程	637	3 676	5	3 676
大学计算机课程	669	3 405	7	3 405

18 331 条边，整合了高中数学、大学数学和大学计算机课程 3 个数据集的知识点，并去除了重复节点和孤立节点，以确保图谱结构的合理性与连通性。其中，高中数学部分提供了基础概念；大学数学进一步扩展了数学理论；计算机课程包含了算法、数据结构等计算机科学知识，与数学知识点形成紧密关联。融合后的知识图谱具有更丰富的层次结构，提高了知识点之间的关联度，增强了知识点之间的信息传播能力。同时，融合后的节点分布更均衡，确保了跨学科知识点在图谱中的合理融合，使其能够更好地支持个性化学习路径推荐和教育资源优化等应用场景。

3.2 实验设置

(1) 参数设置

实验采用 Adam 作为优化器，并针对不同的模型参数进行调优。实验设置 TSFM 的嵌入维度 d_{embed} 为 128；在 GAT 部分，设置注意力头 h_{att} 为 8；设置隐藏层维度 d_{hidden} 为 64。为防止过拟合，Dropout 比例采用 $p_{\text{dropout}} = 0.5$ ，并使用 Xavier 对权重进行初始化。模型的训练参数包括学习率 $\text{lr} = 0.001$ 、训练轮数 $\text{epochs} = 50$ 、批次大小 $\text{batch}_{\text{size}} = 128$ ，并采用 L2 正则化（正则化系数 $\lambda = 0.0001$ ）提升模型的泛化能力。

在计算节点重要性分数时，实验引入语义相似度和结构相似度的加权系数 α 和 β ，以平衡这两种相似度信息对最终评分结果的贡献。经过实验调优，本文发现当 $\alpha = 1$ 、 $\beta = 1$ 时，模型在 NDCG@50 和 Spearman 相关系数上的表现最佳，因此 α 、 β 采用该默认配置。

(2) 基线方法

为了全面评估 TSFM 的有效性，本文选择了两类具有代表性的基准方法进行对

比分析，包括基于统计指标的传统方法和基于图神经网络的可训练方法。

① 基于统计指标的传统方法

基于统计指标的传统方法主要包括 PageRank 和 PPR。PageRank 通过随机游走机制衡量节点的重要性，其核心思想是：一个节点的影响力越大，被其他节点引用的概率就越高。PPR 在 PageRank 的基础上增加了个性化跳转机制，使节点重要性评估更符合特定领域的需求。这两种方法广泛应用于网络分析和搜索引擎排名，但其仅依赖拓扑结构信息，不能充分利用语义信息进行节点重要性评估。

② 基于图神经网络的可训练方法

采用基于图神经网络的可训练方法可以提升节点重要性评估的效果。其中，GCN 通过图卷积操作聚合邻居信息，生成高维嵌入；GAT 进一步引入注意力机制，能够动态调整邻居节点对目标节点的重要性贡献，从而提高嵌入表示的质量；GENI 在 GNN 的基础上融合了中心度指标，使节点重要性评估更加精准；MultiImport 进一步结合了外部输入的重要性信号，增强了模型的适应能力；RGTN 采用 Transformer 结构建模图数据，并结合语义和结构特征，在大规模知识图谱场景下表现较好。

为保证实验的公平性，基准方法的参数设置严格遵循之前各研究学者的设定。PageRank 的阻尼因子 α 设为 0.85，PPR 的跳转概率 p 设为 0.15。对于 GCN 和 GAT，本文统一采用两层结构，隐藏层维度 d_{hidden} 设为 64 学习率 lr 为 0.001，训练轮数 $\text{epochs} = 50$ ，并保持实验条件的一致性。此外，GENI、MultiImport 和 RGTN 分别采用文献[19]、文献[20]、文献[21]的默认参数，以保证实验对比结果的可复现性。

3.3 评价指标

为了综合评估模型在节点排名质量及其与真实重要性之间的一致性,本文采用Huang等^[21]提出的评估指标,包括归一化折损累积增益(normalized discounted cumulative gain, NDCG)、Spearman排名相关系数以及重叠率指标。需要说明的是,仅对那些具备明确真实重要性标签的节点样本进行评估,以确保评价的客观性与有效性。

(1) NDCG

NDCG是评估排名质量的常用指标,能够有效反映排序中元素位置的重要性^[35]。在节点重要性评估任务中,本文将分级相关性值定义为节点真实重要性分数经对数变换后的结果。在获得节点列表、模型预测的节点重要性得分以及对应的真实标签后,模型根据预测得分对节点进行降序排序,并将排序中第*i*位节点的真实得分作为 rel_i 参与 $DCG@k$ ^[21]计算,其定义如下。

$$DCG@k = \sum_{i=1}^k \frac{rel_i}{\log_2(i+1)} \quad (12)$$

通过计算位置*k*处的真实值的DCG获得位置*k*处的理想化折损累积增益(ideal discounted cumulative gain, IDCG),位置*k*处的归一化折损累积增益计算如下。

$$NDCG@k = \frac{DCG@k}{IDCG@k} \quad (13)$$

(2) Spearman排名相关系数

Spearman排名相关系数评估节点重要性列表 l_p 和真实值列表 l_g 之间的相关性^[36]。将 l_p 和 l_g 转换为等级值 R_{l_p} 和 R_{l_g} 。

$$Spearman = \frac{cov(R_{l_p}, R_{l_g})}{\sigma_{R_{l_p}} \sigma_{R_{l_g}}} \quad (14)$$

其中, $cov(R_{l_p}, R_{l_g})$ 是排名变量的协方差;

$\sigma_{R_{l_p}}$ 、 $\sigma_{R_{l_g}}$ 是排名变量的标准差。

(3) 重叠指标

重叠指标是预测重要节点与真实重要节点的重叠率^[37]。本文将预测出的前*k*个重要节点的集合定义为 L'_{top-k} , 将前*k*个真实重要性的节点集合定义为 L_{top-k} 。

$$OVER@k = \frac{|L'_{top-k} \cap L_{top-k}|}{k} \quad (15)$$

3.4 节点重要性评估实验结果

为了确保实验结果的稳定性和可靠性,本文对所有模型进行10次独立重复实验,并以“均值±标准差”的形式呈现最终结果。所有实验均在固定随机种子的条件下进行,以减少实验波动对实验结果的影响,确保实验结果的可复现性。节点重要性评估实验结果见表2。

由表2可知,PageRank和PPR的表现证实了拓扑结构信息的重要性。当引入图神经网络(如GCN和GAT)后,模型的评估效果显著提升。GCN通过聚合邻居节点信息,利用图的结构特征进行节点嵌入,从而提升了节点重要性评估的准确性(NDCG@50为0.8523, OVER@50为0.3845),显示了在信息传播效率上的显著提升。GAT引入了注意力机制,使模型能够动态地为邻居节点分配不同的权重,进一步优化了节点间关系的捕捉与表示, NDCG@50为0.8685, OVER@50为0.4223,显著提升了评估效果。GENI和Multiimport在NDCG@50、Spearman相关系数和OVER@50指标上均有不错的表现, GENI的NDCG@50为0.8715, Spearman相关系数为0.7090, OVER@50为0.4569; MultiImport的NDCG@50为0.8593, Spearman相关系数为0.7111, OVER@50为0.4196。以上结果表明

表2 实验结果(最好的结果加粗显示,次好的结果加下划线)

模型方法	NDCG@50	Spearman相关系数	OVER@50
PR	0.8123 ± 0.007	0.3164 ± 0.015	0.1139 ± 0.011
PPR	0.8217 ± 0.006	0.3173 ± 0.012	0.1263 ± 0.009
GCN	0.8523 ± 0.009	0.6894 ± 0.006	0.3845 ± 0.062
GAT	0.8685 ± 0.005	0.6823 ± 0.011	0.4223 ± 0.056
GENI	0.8715 ± 0.003	0.7090 ± 0.009	<u>0.4569 ± 0.046</u>
Multiimport	0.8593 ± 0.008	0.7111 ± 0.003	0.4196 ± 0.031
RGTN	<u>0.8811 ± 0.005</u>	<u>0.7492 ± 0.011</u>	0.4335 ± 0.063
TSFM	0.9017 ± 0.004	0.7713 ± 0.006	0.4873 ± 0.035

GENI和Multiimport能够有效地聚合邻域信息，并准确评估节点的重要性。尤其是在Spearman相关系数上，GENI和Multiimport表现出的相关性较高，说明了它们在节点排序方面的准确性。RGTN在所有指标上的表现均相对出色，特别是Spearman相关系数达到了0.7492，显示了其在评估节点重要性方面的高度准确性和可靠性。RGTN通过引入两个不同方向的编码器从而获得了更精准的节点嵌入，显著提高了节点重要性评估的准确度。TSFM在所有评价指标上都取得了最佳成绩，尤其是NDCG@50达到了0.9017、OVER@50达到了0.4873，表明该模型通

过整合拓扑结构信息和节点特征，展现了其在捕获节点重要性方面的强大能力。此外，TSFM在Spearman相关系数上的表现同样优异，进一步验证了该模型在知识图谱节点排序任务中的有效性。

3.5 超参数分析实验

α 和 β 分别控制语义相似度和结构相似度在节点重要性计算中的加权重。为了进一步优化节点重要性评估模型的性能，本实验重点分析模型中的这两个关键超参数： α 和 β 。实验旨在研究不同 α 和 β 比值对模型评估结果的影响，并探索最佳的超参数配置。

在该实验中，本文设置了不同的 α 和 β 比值进行实验，具体的比值选择包括0.1、0.5、1、5、10。通过调整这两个超参数的比值，研究其对模型性能的影响。特别地，本文通过对比不同比值下的NDCG@50和Spearman相关系数来评估模型的节点重要性评估效果。实验结果如图3所示。

由图3可知， α 和 β 的比值对模型的性能具有显著影响。当 α 和 β 的比值为0.1时，模型的表现最差。此时，语义相似度

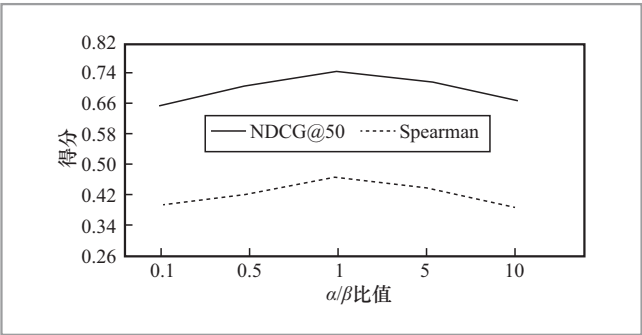


图3 超参数分析实验结果

的权重远低于结构相似度，导致模型过度依赖拓扑结构，忽视了节点间的语义关联，导致 NDCG@50 和 Spearman 相关系数均较低。随着比值增大，模型逐渐开始平衡语义相似度与结构相似度的影响。当 α 和 β 的比值为 1 时，模型表现最佳，NDCG@50 和 Spearman 相关系数均达到了较高值，表明在节点重要性评估中，语义和结构相似度的加权应保持平衡。在该配置下，模型能够较准确地捕捉到节点间的语义和结构关系，确保评估结果的全面性和准确性。当 α 和 β 比值超过 1 时，尤其是比值为 5 和 10 时，模型表现逐渐下降，这是因为过重的语义相似度权重使模型忽视了拓扑结构的重要性，尤其是在处理结构复杂的知识图谱时，模型不能有效捕捉节点间的拓扑关系，导致节点重要性评估精度下降。因此，尽管语义相似度在某些应用场景下具有重要作用，但过度依赖语义信息会对模型的整体表现产生负面影响。

综上所述，超参数分析实验结果表明，合理的 α 和 β 比值对模型性能具有重要影响，特别是语义和结构相似度的平衡在节点重要性评估任务中至关重要。因此，推荐在实际应用中采用 $\alpha/\beta=1$ 的配置作为优化模型性能的默认设置。未来的工作可以进一步探索在不同类型的教育知识图谱中，如何根据具体需求调整超参数，以进一步提升评估精度和模型的适用性。

3.6 消融实验

为了进一步验证 TSFM 中各个组件的重要性，本文进行了消融实验。本文分别移除拓扑结构信息、语义相似度和结构相似度，观察模型性能的变化。同样，所有实验均进行了 10 次重复实验，并计算了标准差。实验结果见表 3。

由表 3 可知，移除拓扑结构信息后，模型在所有评价指标上均有显著下降。特别是在 NDCG@50 指标上下降了约 13.6%，这表明拓扑结构信息对于模型识别重要节点至关重要。同样，当移除语义相似度时，模型的 Spearman 相关系数减少了约 15.1%，表明语义信息对于捕捉节点间深层次关系至关重要。结构相似度的移除对模型影响最大，特别是在 OVER@50 指标上下降了约 32.5%，这表明结构相似度在评估节点在图中的结构作用方面起关键作用。

综上所述，拓扑结构信息、语义相似度和结构相似度共同构成了 TSFM 的核心部分，三者的相互配合显著提升了模型的评估效果。

3.7 案例分析

为了进一步验证 TSFM 在知识图谱节点重要性评估中的有效性，本文选取了典型知识点进行案例分析，比较它们在不同

表3 消融实验结果

模型配置	NDCG@50	Spearman相关系数	OVER@50
移除拓扑结构信息	0.6952 ± 0.004	0.5731 ± 0.002	0.3145 ± 0.023
移除语义相似度	0.7266 ± 0.006	0.6123 ± 0.006	0.3252 ± 0.051
移除结构相似度	0.7134 ± 0.002	0.5856 ± 0.005	0.2908 ± 0.013
TSFM	0.8045 ± 0.007	0.7216 ± 0.003	0.4163 ± 0.012

方法（PageRank、GCN、GAT、RGTN、TSFM）下的节点重要性排名变化。实验旨在展示 TSFM 如何在语义和结构相结合的情况下，更准确地识别关键知识点，并优化节点的重要性排序。

本文从融合后的大型跨学科知识图谱中选取了几个核心知识点，包括线性代数、数据结构、计算机网络、操作系统、机器学习等。这些知识点在不同教学场景下对学习路径推荐、课程规划具有重要影响。在案例分析中，对比这些知识点在 PageRank、GCN、GAT、RGTN 和 TSFM 下的排名。**表 4** 展示了各个知识点在不同方法下的重要性排名（数值越小表示排名越高，即节点越重要）。

由**表 4**可知，PageRank 主要依赖度中心度，导致部分知识点排名远偏离其真实重要性。例如，在真实排名中，“数据结构”“线性代数”分别排第 1 和第 2，但在 PageRank 评估下分别排第 4 和第 5，说明 PageRank 更关注连接度，忽略了知识点在全局学习路径中的核心地位。GCN 评估的“线性代数”排名为第 7，远低于其真实排名（第 2）；GAT 评估的“数据结构”为第 5，不能准确反映其核心作用。这表明，GCN、GAT 虽然增强了局部结构学习能力，但在全局知识传播方面仍存在局限。RGTN 和 TSFM 综合考虑了拓扑结构和节点特征，在全局重要性评估上表现更优。

其中，RGTN 的评估结果已较 PageRank 和 GCN 有所改善，但仍存在偏差，如“数据结构”仅升至第 3，仍未匹配真实排名（第 1）。相比之下，TSFM 评估出的“数据结构”排名最接近真实排名、“计算机网络”和“机器学习”的排名分别为第 4 和第 3，更符合其在计算机科学课程体系中的重要性。这表明，TSFM 能够更精准地识别关键知识点，并在教育知识图谱应用中表现更优。

综合来看，TSFM 在多个核心知识点上的排名更符合真实学习场景，说明其能够精准识别真正重要的知识点。与其他方法相比，TSFM 通过结合拓扑结构和语义信息，避免了 PageRank 依赖度中心性的问题，同时弥补了 GCN 和 GAT 过度依赖局部信息的缺陷，最终评估出的知识点排名与真实重要性排名高度匹配。其证明了 TSFM 在教育知识图谱中的适用性更强，为个性化学习路径推荐和知识点筛选提供了更精准的支持。

4 结束语

本文提出了一种基于拓扑结构信息与重要节点相似度信息相结合的节点重要性评估模型 TSFM，旨在应对教育知识图谱中识别关键节点的挑战。在模型架构方面，

表 4 节点重要性排名对比结果

知识节点	真实排名	Pagerank 预测的排名	GCN 预测的排名	GAT 预测的排名	RGTN 预测的排名	TSFM 预测的排名
线性代数	2	5	7	6	4	3
数据结构	1	4	5	5	3	2
计算机网络	4	8	6	7	5	4
操作系统	5	10	9	8	6	5
机器学习	3	7	8	8	5	3

本文首先计算了节点在知识图谱中的拓扑特征，包括中心度因子、出入度因子和通过率因子，进而优化节点嵌入表示，为评估节点的重要性提供了初始的定量基础；然后，利用 GAT 对节点的邻域信息进行了综合分析并聚合，进一步提高了模型对节点间复杂相互作用的捕捉能力；最后，模型根据节点与重要节点之间的相似度，对初始重要性得分进行加权优化，输出更准确的节点重要性评估结果。

TSFM 为教育知识图谱中的关键节点识别提供了一种更有效的方法，不仅有助于提升对知识点重要性的准确评估，还能够优化教育资源的合理分配。此外，在学习路径推荐方面，TSFM 能够根据学生在知识图谱中的学习进展，智能地引导学生通过合理的路径学习新知识，减少冗余学习过程。TSFM 在个性化教育和智慧教育中具有极大的应用潜力，能够帮助教育者更好地因材施教，并支持学习者的个性化发展。

未来，笔者计划进一步探索将更多类型的信息融入模型，以提高节点重要性评估的准确性和鲁棒性。例如，考虑引入时间维度的信息来处理教育知识图谱中的动态变化，帮助模型适应知识点和关系的更新；结合多模态数据（如文本、视频、课程资源等）进行综合分析，更全面地反映知识点的多层次关联性和实际教学需求，从而提升模型在不同场景中的适用性。

参考文献：

[1] 专题：大模型与知识图谱[J]. 大数据, 2025, 11(2):3-4.
Large language model meets knowledge graph[J]. Big Data Research, 2025, 11(2):

3-4.
[2] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606.
Xu Z L, Sheng Y P, He L R, et al. Review on knowledge graph techniques[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4): 589-606.
[3] 范佳荣, 钟绍春. 学科知识图谱研究：由知识学习走向思维发展[J]. 电化教育研究, 2022, 43(1): 32-38.
Fan J R, Zhong S C. Research on subject knowledge mapping: from knowledge learning to thinking development[J]. e-Education research, 2022, 43(1): 32-38.
[4] 李振, 周东岱. 教育知识图谱的概念模型与构建方法研究[J]. 电化教育研究, 2019, 40(8): 78-86, 113.
Li Z, Zhou D D. Research on conceptual model and construction method of educational knowledge graph[J]. e-Education research, 2019, 40(8): 78-86, 113.
[5] 余胜泉, 李晓庆. 区域性教育大数据总体架构与应用模型[J]. 中国电化教育, 2019(1): 18-27.
Yu S Q, Li X Q. The study on architecture and application model about regional education big data[J]. China Educational Technology, 2019(1): 18-27.
[6] 李艳燕, 张香玲, 李新, 等. 面向智慧教育的学科知识图谱构建与创新应用[J]. 电化教育研究, 2019, 40(8): 60-69.
Li Y Y, Zhang X L, Li X, et al. Construction and innovative application of discipline knowledge graph oriented to smart education[J]. e-Education research, 2019, 40(8): 60-69.
[7] 和文斌, 董永权, 赵成杰, 等. 基于学科知识图谱的教育知识服务模型构建研究[J]. 数字教

- 育, 2022, 8(6): 21–28.
- He W B, Dong Y Q, Zhao C J, et al. Research on the construction of educational knowledge service model based on discipline knowledge graph[J]. Digital Education, 2022, 8(6): 21–28.
- [8] 高茂, 张丽萍. 融合多模态资源的教育知识图谱的内涵、技术与应用研究[J]. 计算机应用研究, 2022, 39(8): 2257–2267.
- Gao M, Zhang L P. Research on connotation, technology and application of educational knowledge graph based on multi-modal resources[J]. Application Research of Computers, 2022, 39(8): 2257–2267.
- [9] Qiong Q, Dong X W. Evaluation method for node importance in complex networks based on eccentricity of node[C]//Proceedings of the 2016 2nd IEEE International Conference on Computer and Communications (ICCC). Chengdu: IEEE, 2016.
- [10] Xu G W. Identify influential nodes in complex networks based on modified TOPSIS[C]//Proceedings of the 2017 36th Chinese Control Conference (CCC). Dalian: IEEE, 2017.
- [11] Ai X B. Node importance ranking of complex networks with entropy variation[J]. Entropy, 2017, 19(7): 303.
- [12] Le L, Hewei Y. A new method for evaluating node importance in complex networks based on data field theory[C]//Proceedings of the 2010 First International Conference on Networking and Distributed Computing. Hangzhou: IEEE, 2010.
- [13] Hu P, Fan W L, Mei S W. Identifying node importance in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2015, 429: 169–176.
- [14] Page L, Brin S, Motwani R, et al. The PageRank citation ranking: Bringing order to the web[C]//Proceedings of the 7th International Conference on World Wide Web. Toronto: ACM, 1999.
- [15] Haveliwala T H. Topic-sensitive pagerank[C]//Proceedings of the 11th International Conference on World Wide Web. Hawaii: ACM, 2002.
- [16] Tong H H, Faloutsos C, Pan J Y. Random walk with restart: fast solutions and applications[J]. Knowledge and Information Systems, 2008, 14(3): 327–346.
- [17] Li X, Ng M K, Ye Y. HAR: hub, authority and relevance scores in multi-relational data for query search [C]//Proceedings of the 2012 SIAM International Conference on Data Mining. California: SIAM, 2012.
- [18] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[EB]. arXiv preprint, 2016, arXiv: 1609.02907.
- [19] Park N, Kan A, Dong X L, et al. Estimating node importance in knowledge graphs using graph neural networks[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Alaska: ACM, 2019.
- [20] Park N, Kan A, Dong X L, et al. Multi-import: inferring node importance in a knowledge graph from multiple input signals[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. San Diego: ACM, 2020.
- [21] Huang H, Sun L, Du B, et al. Representation learning on knowledge graphs for node importance estimation[C]//Pro-

- ceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Singapore: ACM, 2021.
- [22] Grover A, Leskovec J. Node2vec: scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016.
- [23] Velickovic P, Cucurull G, Casanova A, et al. Graph attention networks[EB]. arXiv preprint, 2017, arXiv: 1710.10903.
- [24] Huang C, Fang Y, Lin X, et al. Estimating node importance values in heterogeneous information networks[C]//Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE). Kuala Lumpur: IEEE, 2022.
- [25] Hamilton W L, Ying R, Leskovec J. Inductive representation learning on large graphs[EB]. arXiv preprint, 2017, arXiv: 1706.02216.
- [26] Xu K, Hu W, Leskovec J, et al. "How powerful are graph neural networks?" [C]//Proceedings of the International Conference on Learning Representations (ICLR). New Orleans: Unknow, 2019.
- [27] Zhang C, Song D, Huang C, et al. Heterogeneous graph neural network [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019.
- [28] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks. Springer International Publishing, 2018: 593–607.
- [29] Dijkstra E W. A note on two problems in connexion with graphs[J]. *Numerische Mathematik*, 1959, 1(1): 269–271.
- [30] Sun Z Q, Deng Z H, Nie J Y, et al. RotatE: knowledge graph embedding by relational rotation in complex space[EB]. arXiv preprint, 2019, arXiv: 1902.10197.
- [31] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Lake Tahoe: Neural Information Processing Systems, 2013(2): 2787–2795.
- [32] Yang B S, Yih W T, He X D, et al. Embedding nodes and relations for learning and inference in knowledge bases[EB]. arXiv preprint, 2014, arXiv: 1412.6575.
- [33] Sun Z, Deng Z H, Nie J Y, et al. Rotate: knowledge graph embedding by relational rotation in complex space[J]. arXiv preprint arXiv:1902.10197, 2019.
- [34] Schmidhuber J. Deep learning in neural networks: an overview[J]. *Neural Networks*, 2015, 61: 85–117.
- [35] Järvelin K, Kekäläinen J. Cumulated gain-based evaluation of IR techniques [J]. *ACM Transactions on Information Systems*, 2002, 20(4): 422–446.
- [36] Zar J H. Significance testing of the spearman rank correlation coefficient[J]. *Journal of the American Statistical Association*, 1972, 67(339): 578–580.
- [37] Zhang T Y, Hou C B, Jiang R, et al. Label informed contrastive pretraining for node importance estimation on knowledge graphs[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(3): 4462–4476.

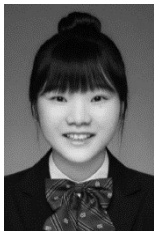
作者简介



李美子（1979-），女，博士，上海师范大学信息与机电工程学院高级实验师，主要研究方向为社交网络分析、大数据分析、人工智能教育等。



伍云芳（2000-），男，上海师范大学信息与机电工程学院硕士生，主要研究方向为知识图谱。



卢淑怡（1998-），女，硕士毕业于上海师范大学信息与机电工程学院，现为华东师范大学教育学部教育信息技术学系博士生，主要研究方向为知识图谱、大模型的教育应用。



王浩（1980-），男，上海新致软件股份有限公司副高级、副总裁，主要研究方向为大数据及人工智能。



杨茹（1992-），女，博士，上海师范大学信息与机电工程学院副教授，主要研究方向为人工智能+教育、形式化方法等。

录用日期: 2025-05-16
通信作者: 杨茹, yangru@shnu.edu.cn
基金项目: 国家自然科学基金项目(No.62477032);国家自然科学基金项目(No.62302306);国家自然科学基金项目(No.62372300)
Foundation Items: The National Natural Science Foundation of China (No.62477032), The National Natural Science Foundation of China (No.62302306), The National Natural Science Foundation of China (No.62372300)