教育数据安全共享中的联邦模型威胁溯源机制与预警策略研究

周俊成* 李兰 彭倩坤 阳城伟

（重庆城市科技学院，重庆 永川 402167）

摘 要：随着教育数字化转型深入推进，跨单位教育数据共享需求日益迫切，联邦学习（Federated Learning，FL）因其“数据不动、模型动”的隐私保护特性已成为教育数据安全共享的重要范式。然而，FL 在教育场景中面临梯度反演、模型投毒及后门攻击等多类安全威胁，现有防御机制在威胁溯源与预警响应方面存在明显不足，尤其缺乏对参与节点行为时序演变规律的动态建模能力。本研究提出融合时序知识图谱的联邦教育模型威胁溯源机制与动态预警策略框架（TKG-FedAlert）。该框架通过构建教育联邦节点行为时序知识图谱，对参与方的梯度更新模式、行为序列及关联关系进行结构化建模；采用时序图注意力网络（TGAT）与时序异常检测相结合的方法实现对恶意节点的精准溯源与威胁预警；并设计三级响应机制支持对潜在威胁的实时干预。基于真实高校联邦学习场景的仿真实验表明，TKG-FedAlert 在威胁溯源准确率（84.2%）、预警 F1 分数（88.4%）及系统响应延迟（1.9 轮次）等核心指标上均显著优于现有基线方法，为构建安全可信的教育联邦数据共享体系提供了理论支撑与实践路径。

关键词：联邦学习；教育数据安全；时序知识图谱；威胁溯源；预警策略；图神经网络

DOI：https://doi.org/10.65196/nma80m18

引言

近年来，教育数字化转型已上升为国家战略。《教育信息化 2.0 行动计划》等政策文件明确提出，要加快教育大数据平台建设，推动跨区域、跨层级教育数据的互联互通与协同应用[14]。在此背景下，高校之间如何在保护学生隐私的前提下实现数据价值的充分挖掘，已成为亟待解决的核心议题。

联邦学习（Federated Learning，FL）由 McMahan 等人于 2017 年系统提出[1]，其核心思想是在不共享原始数据的前提下，通过各参与方本地训练并上传模型梯度或参数，由中心服务器进行聚合迭代，实现数据不出域、模型共建共享的目标。这一特性与教育数据的高隐私性和强监管性高度契合，在自适应学习、学习行为分析及教育质量评估等场景中具有广泛的应用前景。近年来，面向教育领域的联邦学习隐私保护研究进一步揭示了 FL 在满足《个人信息保护法》（PIPL）等合规要求方面的独特优势[16]。

然而，联邦学习并非天然安全。Zhu 等人[2]从理论上证明，仅凭梯度信息即可以较高精度重建训练样本，对包含学生敏感学习记录的教育数据构成严重威胁；Bagdasaryan 等人[3]揭示了比例缩放型后门攻击可绕过主流聚合算法的防御机制；针对联邦学习边缘场景的安全综述[19]进一步系统梳理了恶意梯度注入、协同后门等新型攻击形态。上述威胁在教育场景中的危害尤为严峻，一旦攻击得逞，不仅导致学生隐私数据泄露，还将影响教育决策的科学性与公平性。

当前，针对联邦学习安全的研究主要集中在拜占庭鲁棒聚合、差分隐私和安全多方计算等防

基金项目：重庆城市科技学院 2025 年校级项目《基于时序知识图谱的联邦教育模型安全威胁溯源与预警研究》（项目编号：CKKY2025013）；重庆市高等教育学会高等教育科学研究课题《人工智能赋能教育转型背景下重庆高校大学生人机协同学习的“双维耦合”机制及实证研究》（项目编号：cqgj25208C）。

作者简介：周俊成（1984-），男，硕士研究生，研究方向为人工智能、数据安全。
李 兰（1986-），女，硕士研究生，研究方向为企业管理。
彭倩坤（1988-），女，硕士研究生，研究方向为安全工程。
阳城伟（1996-），男，本科，研究方向为人工智能。
通讯作者：周俊成

御机制上，但在威胁溯源与预警方面存在以下明显不足：一是现有方法多为静态、单轮次分析，缺乏对参与节点行为时序演变规律的动态建模能力；二是溯源粒度较粗，难以区分偶发性异常与持续性恶意行为；三是预警机制滞后，往往在攻击已造成实质性损害后才发出告警；四是现有框架缺乏对教育场景特殊性（数据周期性分布、机构可溯性）的针对性设计。

时序知识图谱（Temporal Knowledge Graph，TKG）能够对实体关系随时间的演变规律进行建模，在复杂行为序列分析方面具有独特优势[6]。面向动态图的深度异常检测方法近年来也取得重要进展[21, 23]，为本研究提供了坚实的方法论支撑。基于上述分析，本研究提出 TKG-FedAlert 框架，旨在通过结构化建模参与节点的行为时序关系，实现对潜在安全威胁的动态识别、精准溯源与实时预警。

相关研究综述

一、联邦学习安全威胁研究

联邦学习的安全威胁可从攻击目标和攻击发起位置两个维度进行分类。从攻击目标看，威胁主要分为三类：针对数据隐私的推理攻击、针对模型可用性的投毒攻击以及针对模型完整性的后门攻击。Zhu 等人[2]证明仅凭梯度信息即可高精度重建训练样本，对含学生敏感记录的教育数据构成严重威胁。Bagdasaryan 等人[3]研究表明，后门攻击者通过比例缩放技术可绕过 FedAvg 等主流聚合算法防御，在教育评分模型中植入定向错误。近期联邦学习安全综述[25]进一步系统梳理了各类攻击类型，并指出在 Non-IID 数据场景下现有防御方案的泛化能力普遍不足[22]，而教育数据的强异质性恰好是该挑战的典型体现。

从攻击发起位置看，客户端攻击因参与方众多、监管困难而更具隐蔽性，尤以高级持续性威胁（APT）最为棘手。现有拜占庭鲁棒聚合方法如 Krum[4]、FLTrust[5]能在一定程度上抵御梯度投毒，但对具有时序关联特征的 APT 识别能力有限[23]。基于图注意力网络的联邦入侵检测（FedGAT）[18]证明了图结构建模对提升跨部门攻击检测精度的有效性，但缺乏专门针对教育场景的适应性设计。

二、知识图谱与时序图在安全领域的应用

知识图谱凭借其对实体关系的结构化表示能力，已在网络安全威胁情报分析领域获得广泛应用[10]。知识图谱与大语言模型的结合进一步提升了威胁描述的语义丰富性，为自动化威胁情报生成提供了新路径[17]。时序知识图谱在静态知识图谱基础上引入时间维度，能够建模实体关系随时间的演变规律，在威胁行为序列分析中具有独特优势[6, 11]。

面向时序数据的图神经网络综述[21]系统梳理了 GNN 在时序异常检测任务中的方法论进展，表明基于时序图的方法在捕获跨时间步依赖关系方面显著优于传统方法。专门针对动态图异常检测的综述研究[23]进一步确认了时序图方法在节点级与子图级异常识别中的突出优势，为本研究设计双层检测策略提供了理论依据。然而，现有研究较少将时序知识图谱专门应用于联邦学习安全分析场景，教育联邦学习特有的周期性行为模式也尚未被充分利用。

三、联邦学习预警与响应机制

针对联邦学习的入侵检测系统（FL-IDS）研究近年取得重要进展。FedGATSage[24]将图注意力网络与 GraphSAGE 相结合，在保持结构感知的同时提升了时序攻击模式检测能力。针对 FL-IDS 的系统综述[23]指出，APT 检测是当前 FL 安全领域的核心挑战。FedDebug 框架[13]通过梯度检查点实现了对恶意更新的事后溯源，但其实时性不足，无法应对高速攻击场景。在教育领域，面向 PIPL 合规的联邦隐私保护研究[16]表明，针对教育场景构建定制化安全方案具有重要的现实价值，而教育联邦场景的威胁溯源与预警研究目前仍属空白，亟需填补。

TKG-FedAlert 框架理论设计

一、框架总体架构

TKG-FedAlert 框架由四个核心模块构成：教育联邦行为数据采集层、时序知识图谱构建层、威胁检测与溯源层以及预警响应层（见表 1）。框架的整体设计逻辑体现为：通过行为数据的持续采集驱动图谱的增量更新，以图谱的时序结构支撑检测模型的特征学习，以检测结果触发分级

响应机制，形成感知—建模—分析—响应的完整安全防护闭环。框架设计遵循最小侵入原则（额外通信开销控制在 5%以内）、隐私保护原则（元数据采集采用差分隐私机制）、可解释性原则（溯源结果以知识图谱路径形式呈现）和分级响应原则（根据威胁等级触发差异化干预）。

表 1 TKG-FedAlert 框架核心模块说明

作用维度	主要功能	关键技术
行为数据采集层	采集梯度统计特征、通信元数据、参与方行为日志	差分隐私、安全多方计算
时序知识图谱构建层	构建节点-行为-时间三元组图谱，动态更新实体关系	TKG 嵌入学习[6, 11]、图模式挖掘
威胁检测与溯源层	异常子图识别、恶意节点溯源、攻击路径重建	TGAT[21]、时序异常检测[23]
预警响应层	威胁等级评估、分级告警推送、自动干预触发	规则引擎、强化学习决策

二、时序知识图谱的构建方法

本研究将教育联邦学习场景中的实体定义为三类：参与节点（P）、梯度更新事件（G）和行为特征（F）。基于此，构建时序知识图谱 TKG = (E, R, T)，其中 E 为实体集合，R 为关系类型集合，T 为时间戳集合。图谱中的时序四元组表示为 (h, r, t, τ)，表示实体 h 在时刻 τ 通过关系 r 与实体 t 产生关联，继承了时序知识图谱嵌入学习领域的规范表示体系[6, 11]，并针对联邦学习梯度通信特性进行了场景化扩展。

关系类型 R 涵盖四类核心关系：（1）上传关系（uploads），描述参与节点在特定时间上传梯度更新的行为；（2）相似关系（similar_to），描述不同参与节点梯度更新向量的余弦相似性；（3）异常关系（anomalous_at），描述梯度统计特征偏离历史基线的异常事件；（4）协同关系（colludes_with），描述可能存在共谋攻击的节点对关系。此外，图谱中额外引入教育周期特征嵌入，将学期节律、考试时间节点等教育日历信息编码为时间位置向量，以提升正常周期性波动与真实异常的区分精度。图谱动态更新遵循滑动窗口机制，维护长度为 W 的历史窗口，超出范围的历史数据通过时间衰减权重进行软删除，动态图异常检测领域研究[23]证实了这种增量更新策略在平衡历史记忆与实时响应方面的有效性。

三、基于图神经网络的威胁检测方法

在威胁检测层，本研究采用时序图注意力网络（TGAT）对时序知识图谱进行节点表示学习。TGAT 通过时间编码函数将时间戳映射为连续向量，并与实体特征进行拼接，从而捕获节点行为的时序演变规律，节点 i 在时刻 τ 的表示形式化为：$h_i(\tau) = \mathrm{AGGREGATE}(\{h_j(\tau_j) \oplus \Phi(\tau - \tau_j) \mid (j, \tau_j) \in N(i, \tau)\})$，其中 $\Phi(\cdot)$ 为时间编码函数，$N(i, \tau)$ 为节点 i 在时刻 τ 之前的历史邻居集合。GNN 在时序数据异常检测中的综合表现已获系统性验证[21]，其在捕获空间结构依赖与时间演变特征方面的优势为本研究提供了坚实方法论依据。

本研究设计双层异常检测策略：第一层为节点级异常检测，通过比较当前轮次节点表示与历史基线表示的马氏距离偏差，识别单节点的异常梯度上传行为；第二层为子图级异常检测，通过图模式匹配识别多节点协同的共谋攻击行为[23, 24]。两层结果经加权融合后输出综合异常评分，超过 CUSUM 控制图自适应动态阈值的节点被标记为潜在威胁，以应对教育场景中因数据周期性变化引起的梯度基线漂移。

四、威胁溯源与攻击路径重建

威胁溯源是本框架的核心创新之一。当检测到异常节点后，系统通过在时序知识图谱上执行反向路径搜索，重建攻击的时序传播路径。具体采用基于图的溯源算法（GPA），该算法以异常节点为起点，沿上传关系的逆向链路追踪梯度毒化源头，并结合协同关系子图识别可能存在的分布式协同攻击模式，形式化约束为路径的时序单调性（$\tau_1 < \tau_2 < \ldots < \tau_k$）。溯源结果以可解释的知识图谱路径形式呈现，并通过图谱子结构匹配实现对攻击类型（投毒攻击、后门攻击、

梯度反演等）的自动分类[10, 17]。

五、分级预警与响应策略

预警响应层采用三级威胁分级体系：（1）一级（低威胁，异常评分 0.3—0.6）触发监控加强响应，梯度采样频率提升至正常水平的两倍；（2）二级（中威胁，异常评分 0.6—0.85）触发隔离观察响应，将该节点梯度排除在当前轮次聚合之外并推送管理员审查通知；（3）三级（高威胁，异常评分>0.85）触发紧急处置响应，永久剔除该节点并触发全局模型回滚至最近安全检查点，同步生成完整溯源证据报告。威胁等级综合考量异常评分、溯源置信度和历史违规记录三维度，通过规则引擎自动判定，并支持人工干预覆盖，符合教育管理实践对决策透明性与问责制的要求。

实验设计与结果分析

一、实验环境与数据集设置

本研究基于高校联邦学习场景构建仿真实验环境，模拟 20 个参与高校节点，采用 Non-IID Dirichlet 分布（α=0.5）数据划分策略，模拟真实各校数据异质性。基础数据集包括：教育行为数据集（基于公开 OULAD 开放大学学习分析数据集的重采样版本）和课程成绩预测数据集（来源于某省级教育数字化平台的脱敏数据，已通过伦理审查）。实验环境部署于配备 4 块 NVIDIA A100 GPU（40GB 显存）的服务器集群，PyTorch 2.0 与 PyTorch Geometric 框架支撑图神经网络训练。攻击场景设置：随机选取 3—5 个参与节点作为恶意节点，分别模拟标签翻转投毒攻击、尺度放大后门攻击和协同梯度共谋攻击三种典型威胁，恶意节点比例在 15%—25%之间动态调整。联邦训练设置为 100 轮，图谱滑动窗口长度 W=20，TGAT 模型采用 4 头注意力机制，嵌入维度 128。

二、对比基线方法

本研究选取四种基线方法进行对比：（1）Krum[4]，拜占庭鲁棒聚合方法；（2）FedAvg+IDS，标准联邦平均聚合叠加统计阈值入侵检测系统；（3）FLTrust[5]，通过服务器端根数据集对客户端梯度进行信任评分过滤；（4）FedDebug（简化版）[13]，基于梯度检查点的事后溯源方法。上述基线覆盖了拜占庭鲁棒聚合、统计检测和事后溯源三类主流防御范式，确保对比实验的全面性与代表性。所有实验结果均为 5 次独立运行的均值，95%置信区间通过 bootstrap 方法估计。

三、核心指标评估结果

表 2　各方法核心性能指标对比（%，延迟单位：轮次）

方法	Precision	Recall	F1	溯源准确率	响应延迟	模型准确率
Krum[4]	71.3	63.2	67.0	—	—	82.4
FedAvg+IDS	68.5	70.1	69.3	—	5.2	84.1
FLTrust[5]	76.8	69.4	72.9	—	3.8	85.3
FedDebug[13]	74.2	71.8	72.9	61.5	8.6	83.7
TKG-FedAlert（本研究）	**89.6**	**87.3**	**88.4**	**84.2**	**1.9**	**86.8**

由表 2 可见，TKG-FedAlert 在威胁检测 F1 分数（88.4%）、溯源准确率（84.2%）和响应延迟（1.9 轮）三项关键指标上均显著优于四种基线方法（$p<0.01$，配对 t 检验）。与性能最接近的 FLTrust 相比，F1 分数提升约 15.5 个百分点，响应延迟缩短约 50%。这一优势的核心来源在于时序知识图谱对跨轮次行为演变规律的有效建模，使系统能够在攻击充分暴露之前实现早期预警。值得注意的是，仅 FedDebug 和本研究框架能够输出溯源结果，而多数现有方法仅能实现 201c 能否检测 201d，无法回答 201c 谁是攻击者 201d 这一关键溯源问题。

四、消融实验分析

为验证各模块的贡献，本研究进行系统性消融实验，依次移除时序建模组件（-TKG）、子图

级检测组件（-SubGraph）和分级响应组件（-Grad），结果如表 3 所示。

表 3 消融实验结果

消融变体	F1 分数（%）	溯源准确率（%）	响应延迟（轮次）
TKG-FedAlert（完整版）	**88.4**	**84.2**	**1.9**
移除时序建模（-TKG）	76.1	68.3	2.7
移除子图检测（-SubGraph）	81.3	79.5	2.1
移除分级响应（-Grad）	88.1	84.0	4.3

消融实验结果揭示了各模块的差异化贡献：时序建模组件（TKG）对整体性能贡献最大，移除后 F1 分数下降 12.3 个百分点、溯源准确率下降 15.9 个百分点，印证了动态图异常检测领域关于时序建模不可替代性的理论共识[21, 23]；子图级检测的移除导致溯源准确率下降 4.7 个百分点，说明协同攻击的多节点关联特征是提升溯源精度的关键信号；分级响应组件的移除对检测精度影响可忽略（仅 0.3 个百分点），但响应延迟从 1.9 轮增至 4.3 轮，增幅达 126%，直接证明了其在提升系统实时性方面的不可或缺性。

五、教育场景特殊性分析

与通用联邦学习场景相比，教育场景具有以下特殊性，已通过针对性实验得到验证。其一，数据异质性更强。本框架通过为每个参与节点构建个性化历史基线，在 Non-IID（α=0.5）场景下相比全局统一阈值使 F1 分数提升了 6.8 个百分点，充分验证了针对教育数据异质性专项设计的价值[16]。其二，参与节点可信度相对可溯。高校作为实名认证机构，其行为具有较强可追责性，为基于节点身份的知识图谱溯源提供了合理假设前提。其三，数据更新具有周期性。本框架在时序知识图谱中引入教育周期特征嵌入，将正常周期性波动误报率从无嵌入条件下的 18.3%降至 5.7%，证实了领域先验知识融入是提升教育场景检测精度的关键。

结论

本研究提出的 TKG-FedAlert 框架通过构建教育联邦节点行为的时序结构化图谱，结合 TGAT 实现跨轮次行为建模与双层异常检测，并通过 GPA 算法与三级响应机制完成威胁的精准定位与实时干预。仿真实验系统评估表明，本框架在威胁检测 F1（88.4%）、溯源准确率（84.2%）和响应延迟（1.9 轮）等主要指标上显著优于现有基线方法，且对教育场景特殊性（数据异质性、周期性行为模式、机构可信度）具有良好适应性。在政策与实践层面，本研究可帮助高校信息化管理部门，在推进跨机构联邦学习平台建设的同时，应同步完善安全监测与威胁响应能力，将联邦模型安全治理纳入教育数字化标准规范体系，以保障教育数字化转型的高质量推进。

参考文献：

[1] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS). PMLR, 2017: 1273-1282.

[2] Zhu L, Liu Z, Han S. Deep leakage from gradients[J]. Advances in Neural Information Processing Systems, 2019, 32: 14747-14756.

[3] Bagdasaryan E, Veit A, Hua Y, et al. How to backdoor federated learning[C]//Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics. PMLR, 2020: 2938-2948.

[4] Blanchard P, El Mhamdi E M, Guerraoui R, et al. Machine learning with adversaries: Byzantine tolerant gradient descent[J]. Advances in Neural Information Processing Systems, 2017, 30: 119-129.

[5] Cao X, Fang M, Liu J, et al. FLTrust: Byzantine-robust federated learning via trust bootstrapping[C]//Proceedings of the 28th Network and Distributed System Security Symposium (NDSS). 2022.

[6] Trivedi R, Farajtabar M, Biswal P, et al. DyRep: Learning representations over dynamic graphs[C]//International

Conference on Learning Representations (ICLR). 2019.
[7] Xu Z, Chen Y, Bellet A, et al. A survey of federated learning for edge intelligence[J]. IEEE Communications Surveys & Tutorials, 2021, 23(4): 2391-2424.
[8] 祝智庭, 胡姣. 教育数字化转型的实践逻辑与发展机遇[J]. 电化教育研究, 2022, 43(1): 5-15.
[9] 王运武, 于长虹. 教育数据安全治理：框架构建与路径选择[J]. 中国电化教育, 2023(3): 1-9.
[10] Peng C, Feng X, Hu P, et al. Investigation of threats identification for cybersecurity knowledge graph[C]//2019 IEEE 4th International Conference on Computer and Communication Systems. IEEE, 2019: 117-121.
[11] 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.
[12] Nguyen G H, Lee J B, Rossi R A, et al. Continuous-time dynamic network embeddings[C]//Companion Proceedings of the Web Conference 2018. 2018: 969-976.
[13] Fu C, Li Q, Shen M, et al. FedDebug: Systematic debugging for federated learning applications[C]//Proceedings of the 44th International Conference on Software Engineering (ICSE). IEEE, 2022: 703-715.
[14] 杨宗凯, 吴砥, 郑旭东. 教育信息化 2.0：新时代信息技术变革教育的关键历史跨越[J]. 教育研究, 2018, 39(4): 16-22.
[15] Yin H, Mallya A, Vahdat A, et al. See through gradients: Image batch recovery via GradInversion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 16337-16346.
[16] Chen J, Wang L, Zhang M, et al. Federated learning for teacher data privacy protection: a study in the context of the PIPL[J]. Frontiers in Big Data, 2026. doi:10.3389/fdata.2026.1681382.
[17] Zhang Y, Li Q, Chen H, et al. Improving network threat detection by knowledge graph, large language model, and imbalanced learning[J]. arXiv preprint arXiv:2501.16393, 2025.
[18] Jianping W, Guangqiu Q, Chunming W, et al. Federated learning for network attack detection using attention-based graph neural networks[J]. Scientific Reports, 2024, 14: 19088.
[19] Zhou X, Liang W, Kevin I, et al. Survey: federated learning data security and privacy-preserving in edge-Internet of Things[J]. Artificial Intelligence Review, 2024. doi:10.1007/s10462-024-10774-7.
[20] Choudhury O, Ng M, Mukherjee A, et al. An overview of implementing security and privacy in federated learning[J]. Artificial Intelligence Review, 2024. doi:10.1007/s10462-024-10846-8.
[21] Jin M, Koh H Y, Wen Q, et al. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024. arXiv:2307.03759.
[22] Ma J, Deng M, Shao J, et al. A survey of security threats in federated learning[J]. Complex & Intelligent Systems, 2024. doi:10.1007/s40747-024-01664-0.
[23] Liu K, Dou Y, Zhao Y, et al. Anomaly detection in dynamic graphs: A comprehensive survey[J]. ACM Transactions on Knowledge Discovery from Data, 2024. doi:10.1145/3669906.
[24] Wu Z, Shao Y, Pan L, et al. Graph-based federated learning approach for intrusion detection in IoT networks[J]. Scientific Reports, 2025. doi:10.1038/s41598-025-25175-1.
[25] Li Z, Wu J, Zou X, et al. Advancements in securing federated learning with IDS: a comprehensive review of neural networks and feature engineering techniques for malicious client detection[J]. Artificial Intelligence Review, 2025. doi:10.1007/s10462-024-11082-w.

Research on Threat Attribution Mechanisms and Early Warning Strategies for Federated Models in Secure Educational Data Sharing

ZHOU Juncheng*, LI Lan, PENG Qiankun, YANG Chengwei

(Chongqing Metropolitan College of Sclence and Technology, Chongqing 402167, China)

Abstract: As educational digital transformation accelerates, the demand for cross-institutional educational data sharing has become increasingly urgent. Federated learning (FL), characterized by its privacy-preserving paradigm of keeping data local while sharing model updates, has emerged as a critical framework for secure educational data sharing. However, FL in educational settings faces multifaceted security threats including gradient inversion, model poisoning, and backdoor attacks, while existing defense mechanisms exhibit significant shortcomings in threat traceability and early-warning response—particularly in dynamic modeling of temporal behavioral evolution of participating nodes. This paper proposes TKG-FedAlert, a threat traceability mechanism and dynamic early-warning strategy framework integrating temporal knowledge graphs (TKG) into federated educational model security. The framework constructs a temporal knowledge graph of educational federated node behavioral patterns, enabling structured modeling of gradient update dynamics, behavioral sequences, and inter-node relationships. A temporal graph attention network (TGAT) combined with temporal anomaly detection achieves precise malicious node traceability and threat warning, while a tiered response mechanism supports real-time intervention. Simulation experiments on realistic multi-university FL scenarios demonstrate that TKG-FedAlert significantly outperforms existing baseline methods in threat traceability accuracy (84.2%), early-warning F1 score (88.4%), and system response latency (1.9 rounds), providing both theoretical foundations and practical pathways for building trusted federated educational data sharing ecosystems.

Keywords: federated learning; educational data security; temporal knowledge graph; threat traceability; early-warning strategy; graph neural network