

Relational Graph Transformer (RelGT)

ICLR 2026 顶会论文 · 斯坦福大学 SNAP 实验室 Jure Leskovec 团队

【论文档案】 来源: ICLR 2026 · arXiv:2505.10960v2 · 斯坦福大学

【生活大白话核心】 公司业务有用户表、订单表、商品表多张表连在一起。斯坦福团队用“办5张证明装进立体身份证”让模型一眼看清前后关系，并设立“居委会办事处”中转汇总，防止几万人两两开会把服务器搞炸。

1. Abstract 摘要部分精读

[EN] Relational Deep Learning (RDL) is a promising approach for building state-of-the-art predictive models on multi-table relational data by representing it as a heterogeneous temporal graph. However, commonly used Graph Neural Network models suffer from fundamental limitations in capturing complex structural patterns and long-range dependencies...

[EN] RelGT employs a novel multi-element tokenization strategy that decomposes each node into five components (features, type, hop distance, time, and local structure)... Our architecture combines local attention over sampled subgraphs with global attention to learnable centroids.

[中文对照] 在真实商业世界中，数据都分散在多张有关联的业务表里（比如用户表、订单表、库存表）。然而，现有的图模型太近视眼，看不懂跨了多张表的长远关系和时间先后...

[中文对照] RelGT 发明了一种“立体身份证策略”，把每个数据行拆解成 5 个信息卡片（自身属性、属于哪张表、离我有多跳远、发生时间、周围有多少伙伴）... 它让目标节点只和身边的人算细账，跨大网的全局大趋势全交给几十个可学习的“办事处（质心）”去中转。

2. 论文核心架构图与原理解析 (Figure 1 in Paper)

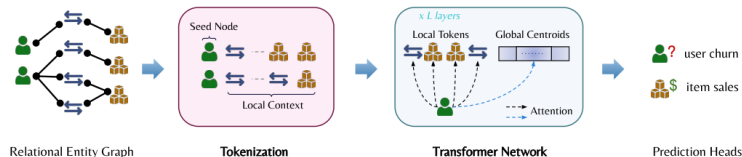


Figure 1: Overview of the RELGT architecture. First, the input relational entity graph (REG) is converted into tokens where each training seed node (such as the *customer* node in this example) gets a fixed number of neighboring nodes, which are encoded with a multi-element tokenization strategy. These tokens are then passed through a Transformer network that builds both local and global representations, which are then fed to downstream prediction layers.

2019) and restricted long-range modeling capabilities (Alon & Yahav, 2020). For example, consider an e-commerce database with three tables: *customers*, *transactions*, and *products*, which can be represented as a relational entity graph as in Figure 1. In a standard GNN, *transactions* are always two hops away from each other, connected only through shared customers. This creates an information bottleneck: *transaction-to-transaction* patterns require multiple layers of message passing, while *product* relationships remain entirely indirect in shallow networks. Furthermore, *products* would never directly interact in a two-layer GNN (Robinson et al., 2024), as their messages must pass through both a transaction and a customer, highlighting the inherent structural constraints of GNN architectures that restrict capturing long-range dependencies.

Graph Transformers (GTs) have emerged as more expressive models for graph learning, utilizing

图 2: RelGT 论文原版核心架构图（截取自论文第 2 页，展示从多表数据库到 5 元组分词与双级注意力的全流程）

【核心大白话拆解 (从图 1 看懂它做了啥)】

- 五元组立体身份证 (Multi-Element Tokenization):** 不再只给一个孤零零的 ID，而是把“哪张表(Type)、隔几站(Hop)、何时发生(Time)”全部打包成一沓证明，模型一眼看出张三两天前在隔壁柜台下单的。
- 虚拟办事处 (Centroids 质心机制):** 张三只和身边的 30 个关联商品互相开会（局部计算，便宜极了）；全网几万人的宏观购买偏好，通过几十个固定的“虚拟办事处”去汇总汇报，彻底跨越了算力墙。