Learning from Relational Structure: Understanding When and How Databases Benefit Predictive Models

Tamara Cucumides, Floris Geerts

Adrem Data Lab, University of Antwerp, Belgium
{tamara.cucumidesfaundez,floris.geerts}@uantwerp.be

Most real-world structured data resides in relational databases, yet the majority of machine learning models for tabular data assume a single-table, row-local, i.i.d. setting. This assumption collapses multi-table relational structure and prevents models from leveraging dependencies that span rows or tables. As a result, traditional tabular models cannot respond to changes elsewhere in the database and fail on tasks whose labels depend on global or relational properties such as counting, uniqueness, or multi-table interactions [1].

A natural alternative is to transform relational data into graphs and apply graph-based learning methods such as graph neural networks or graph transformers [2]. However, current approaches commonly rely on schema-derived graph constructions, which may not expose the structural signals relevant to the predictive task and often disallow informative inter-table connections. Consequently, these graph topologies may be suboptimal for learning.

We introduce *auGraph* [3], a task-aware method for constructing graphs from relational databases. Given a target task, auGraph automatically selects informative attributes and promotes them to graph nodes, enabling message passing through relational pathways that are directly pertinent to prediction. This produces parsimonious, task-relevant graph structures that avoid both underspecification and structural overload.

Our conceptual and empirical analyses reveal a fundamental gap between row-local tabular models and graph-based models, even for simple relational tasks. We further show that auGraph outperforms existing graph-construction baselines, including those that expose all available structure, highlighting the importance of selecting relational structure rather than merely accumulating it. We conclude with implications for scaling to large, noisy real-world databases, developing a theoretical account of relational task structure, and characterizing the expressiveness of different graph-construction paradigms.

References

[1] Tamara Cucumides and Floris Geerts. The missing structure: When graph representations outperform tabular models. In *EurIPS 2025 Workshop: AI for Tabular Data*, 2025.

[2] Matthias Fey, Weihua Hu, Kexin Huang, Jan Eric Lenssen, Rishabh Ranjan, Joshua Robinson, Rex Ying, Jiaxuan You, and Jure Leskovec. Position: Relational deep learning - graph representation learning on relational databases. In *Forty-first International Conference on Machine Learning*, 2024.

[3] Tamara Cucumides and Floris Geerts. From features to structure: Task-aware graph construction for relational and tabular learning with GNNs. In *Proceedings of the VLDB 2025 Workshops: Tabular Data Analysis (TaDA)*. VLDB Endowment, 2025.