PRJ_4SD06_TP • Projet

Beyond Flat Tables: Does Structure Matter?

Anatomy of the Relational Advantage in Relational Deep Learning

ZHENG Tong • NARANJO RODRIGUEZ Natalia • SOUSSOU Koffi Désiré
CIZERON Jules • RYSZFELD Sacha

Instructor **Nikola Simidjievski** • Télécom Paris • RelBench • **June 2026**

The Question & Hypotheses

Do GNNs on relational databases beat flat tabular models?

Many real-world ML problems are not stored as one flat matrix. Information is distributed across multiple linked tables (users, items, events, transactions, races, drivers, results). Flattening requires manual feature engineering, aggregation choices, and a fixed horizon. We test whether models that preserve relational structure can exploit signals difficult to capture through entity-table-only features.

Hypotheses

H1 Relational structure improves predictions

H2 The benefit is *task-dependent*

H3 Flat baselines stay strong

H4 If it helps, the gain is *localizable*

Datasets & Experimental Design

Dataset	Task	T.	Tbl	Signal type
rel-f1	driver-dnf / top3	clf	9	inter-entity
rel-f1	driver-position	reg	9	race-day noise
rel-amazon	user / item-churn	clf	3	activity features
rel-amazon	review-rating	reg	3	interaction
rel-hm	item-sales	reg	3	autoregressive

Temporal RelBench splits — no future leakage

Clf: ROC-AUC / AP-AUPRC / R@10 where relevant

Reg: MAE / RMSE / R^2

Model families

Flat: LightGBM, XGBoost

+ RandomForest / CatBoost on rel-amazon

Relational: heterogeneous GNNs

GraphSAGE, HeteroGraphSAGE, R-GCN / HeteroSAGE

Temporal or sampled neighbours; past-only information

Designed sweep

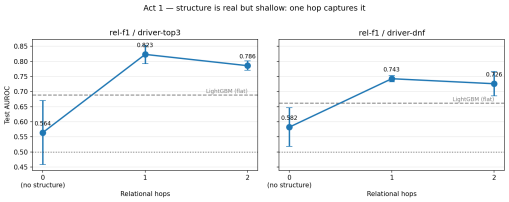
3 databases • 7 tasks

classification + regression

feature-rich vs low-feature regimes

0-hop controls + graph ablations

H1 Confirmed — When structure is the target



Depth ablation — same GNN, vary hops only.

0 hop = near-chance; 1 hop = full gain; 2nd hop adds noise.

Sparse & shallow 1 hop, 1 table, 1 layer.

When the target *is* an entity interaction, topology dominates engineered features.

`rel-f1` classification

Model	top3	dnf
LightGBM (flat)	0.689	0.662
GNN (1 hop)	**0.842**	**0.731**
Full GNN (9 tbl, 2 hop)	0.786	0.726

`rel-amazon` review-rating

Model	MAE
LightGBM (flat)	0.770
No-graph (MLP)	0.641
SAGE 1-layer	**0.592**

H1 & H4 ZHENG Tong, SOUSSOU Koffi

H2 Task dependent behaviour of structure

Same model, different regimes

Task (metric)	Flat	GNN
rel-f1 top3 (AUROC)	0.689	**0.842**
driver-position (MAE)	4.138	$\approx$ 4.128
item-churn (no feat, AUROC)	0.742	**0.777**
item-churn (feat, AUROC)	**0.787**	0.788
item-sales (R^2)	**0.613**	0.563

AUROC $\uparrow$ better MAE $\downarrow$ better R^2 $\uparrow$ better

Same model family $\rightarrow$ completely different outcomes

Key insight

Performance depends on the **type of signal**, not the model

Helps

Interaction-driven targets (rel-f1, review-rating)

Negligible

Features already capture signal (churn, driver-position)

Harmful

Autoregressive signal dominates (item-sales)

Same architecture $\rightarrow$ different regimes $\rightarrow$ H2 confirmed

H3 — Flat baselines remain strong

Same model family, strong tabular baselines

Task (metric)	Flat	GNN
rel-f1 top3 (AUROC)	0.689	**0.842**
driver-position (MAE)	4.138	$\approx$ 4.128
item-churn (no feat, AUROC)	0.742	**0.777**
item-churn (feat, AUROC)	0.787	$\approx$ 0.788
item-sales (R^2)	**0.613**	0.575

AUROC ↑ better MAE ↓ better R^2 ↑ better

Flat models remain competitive across regimes

Key insight

Strong tabular models remain competitive baselines

Competitive

Close to GNN performance (churn with features)

Equivalent

No gain from structure (driver-position)

Superior

Outperforms GNN (item-sales)

Structure is not always needed → H3 confirmed

H4 — Advantage is localizable

Minimal sub-graph vs. full database

rel-f1 top3	Tbl / hop	AUROC
Flat baseline	1 / 0	0.689
Minimal RDL	2 / 1	**0.842**
Full GNN	9 / 2	0.786

AUROC ↑ better — one linked table at one hop suffices

Same shape: review-rating peaks at 1 layer; item-sales at 0 hops

Key insight

The gain lives in a **small sub-schema** — not deep, whole-database message passing

Sparse

Any 1 of 3 linked tables ≈ full 9-table GNN

Shallow

1 hop captures it; the 2nd hop adds noise

Extreme

When harmful, the entity table alone wins (item-sales)

Minimal model ≈ full database → H4 confirmed

Synthesis — When does structure matter?

HELPS	NEGLIGIBLE	HARMFUL
rel-f1 classification review-rating	driver-position churn (with features)	item-sales
Target is an interaction between entities	*Features already capture the signal*	*Autoregressive signal: neighbourhood = noise*

Two cross-cutting regularities

Sparse & shallow: 1 table, 1 hop recovers the full advantage — deeper message passing over-smooths and adds variance.

Features ≈ structure: temporal aggregates replicate neighbourhood aggregation; where they can, flat models stay competitive.

Practical recipe: run 0-hop first. Only pay the graph cost when the target is a cross-entity interaction.

Thank you

Questions?