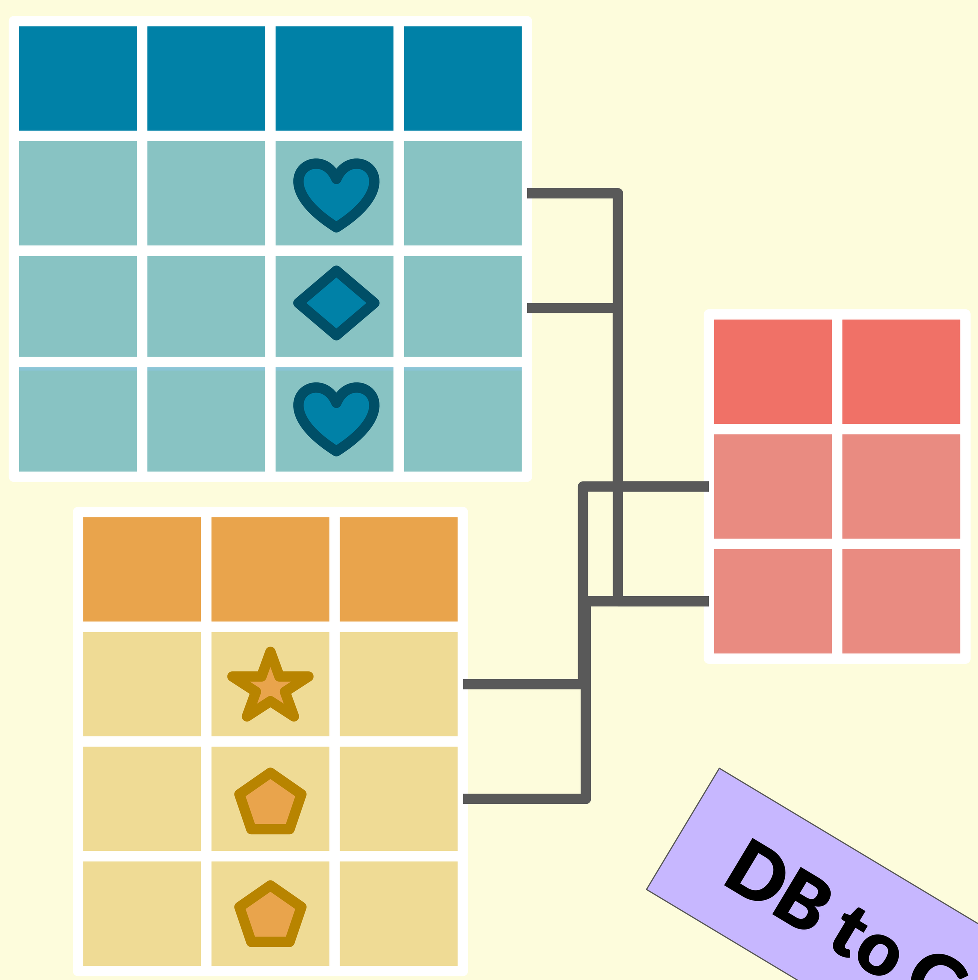
Learning from Relational Structure

Understanding **When and How** Databases Benefit Predictive Models.

Tamara Cucumides, Floris Geerts.

Why should databases matter for machine learning?

- Most data lives in relational databases, no flat tables.
- Traditional ML for tabular data assumes a single-table, single-row view:

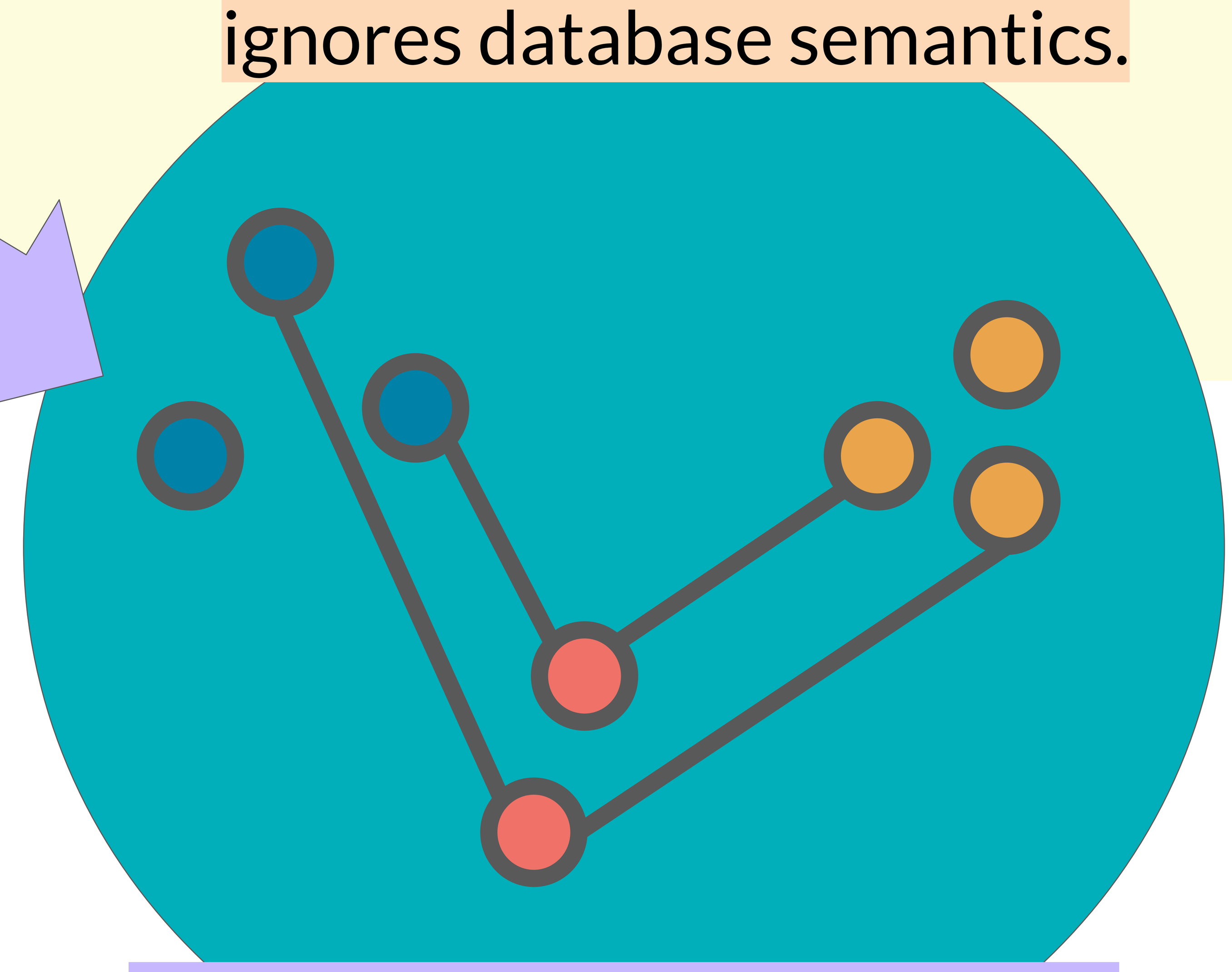


The consequence: **structure is lost.**
Row-local ML cannot:

- React to added or removed rows
- Detect duplicates or shared values
- Compute relational signals (uniqueness, counting, existence)

Structural limitation: row-local ML ignores database semantics.

DB to Graph



Learning in the graph: with a GNN or a Graph Transformer

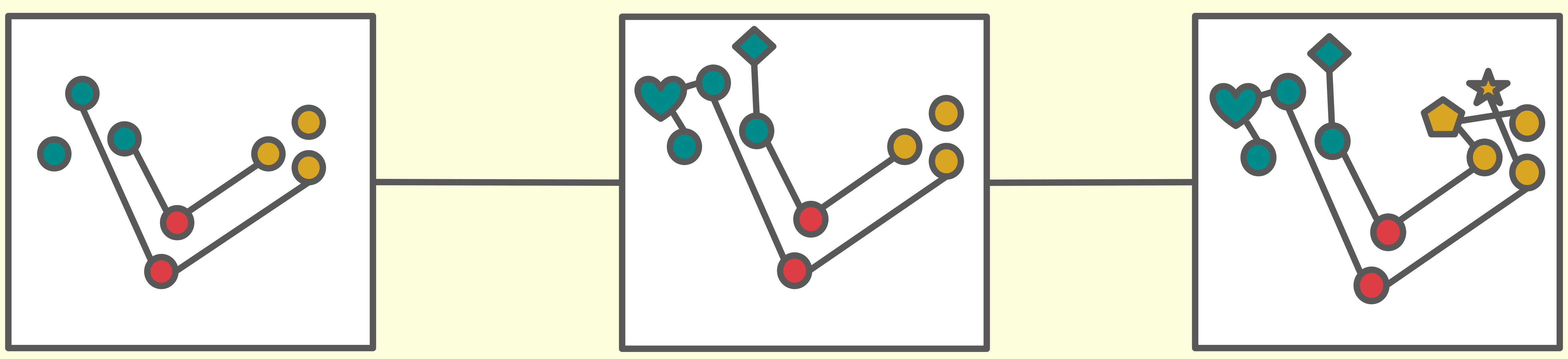
From tables to graphs: GRABLES!

- A natural idea: turn relational data into graphs and apply graph learning. This allows models to:
- Represent rows as nodes, connect them via shared values or key-based links and propagate information using GNNs or transformers

But a database can become many graphs: the challenge is finding the one that helps learning: and auGraph builds it.

Relational entity graph

Augmented graph



The relational entity graph is **successively augmented** by adding new *attribute-nodes* and edges chosen by a **task-aware metric**

Experimental results

1. the structure gap is real: row local models cannot learn uniqueness, and counting
2. auGraph's are better than baselines graphs: better than minimum and maximum added structure

Conclusions and future work

- **Relational structure is needed for many predictive tasks**
- **Graph construction from relational DB is itself a learning step**
- **Databases already contain useful information for learning: we just need to expose it!**

> we need to scale to noisy and real-world DBs
> explore theoretical boundaries: task characterization and expresiveness

Check out our papers!