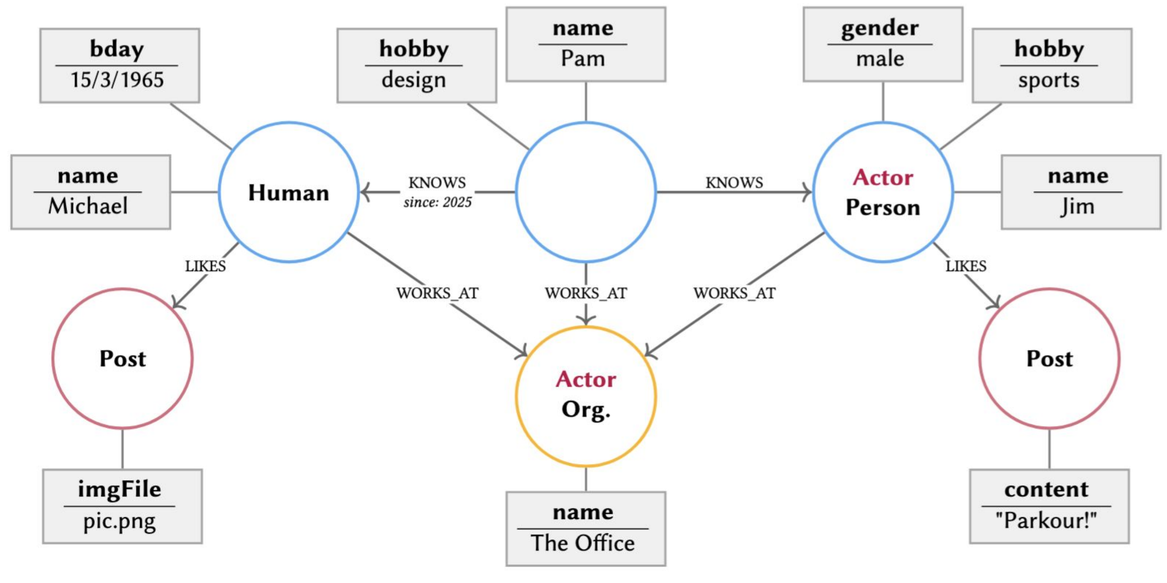
PG-HIVE: Schema Discovery for Property Graphs

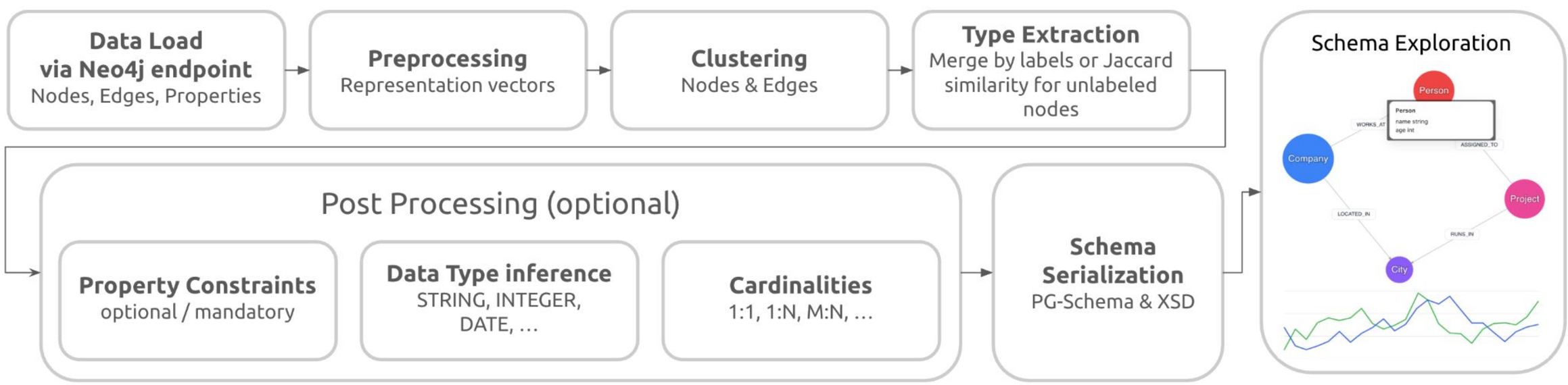
S. Sideri[2,5], I. Chiras[2], M. Giakoumakis[1,2], G. Troullinou[1,4], E. Ymeralli[1,2], V. Efthymiou[3,1], D. Plexousakis[1,2], H. Kondylakis[1,2]
[1]ICS - FORTH, [2]CSD UoC , [3]DIT HUA, [4]CNRS, Univ. Grenoble Alpes, [5]LIPADE, UPCité

Motivation / Problem

- **Property graphs** are **flexible** & **schema-free**
- Missing or inconsistent labels make schema discovery **difficult**
- Existing methods **struggle** with **noisy** and **incomplete** graphs
- Schema discovery is needed for **integration**, **exploration**, **validation**, and **querying**

*"Given a property graph of **arbitrary size and structure**, with **missing type** information, **heterogeneous properties**, and **frequent updates**, infer the schema graph efficiently and accurately."*





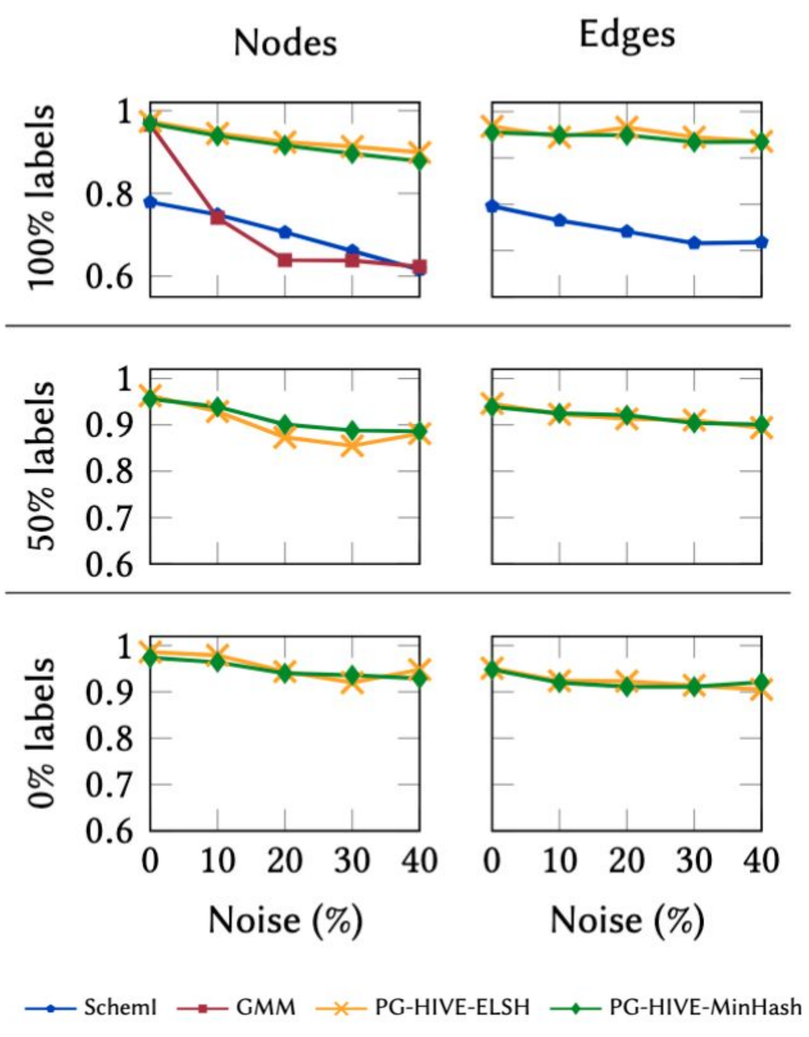
Contributions

- Hybrid schema discovery using **labels**, **properties**, and **graph characteristics**
- Handles missing labels through **structural similarity**
- Supports **incremental updates** without full recomputation
- Infers **schema details**: node & edge types, property data types, mandatory/optional properties, cardinalities

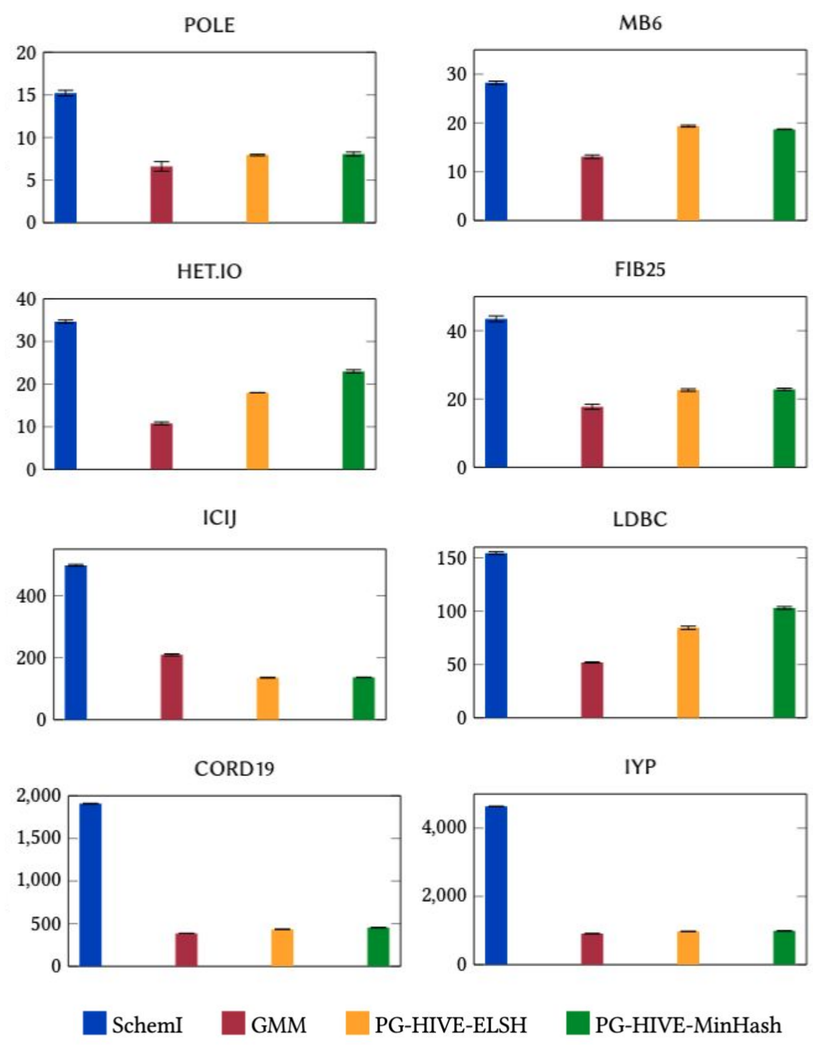
Experiments

- Compared with GMM and SchemI on **synthetic** and **real** datasets.
- To evaluate harder cases, we **add noise** by randomly removing 0%–40% of properties, and test three **label availability scenarios**: 100% labels retained, 50% retained, and 0% labels
- PG-HIVE **outperforms** baseline approaches and infers the schema **successfully** even with noisy and semi-annotated data

Avg F1 Scores across datasets



Type discovery execution time (s)



Results

Accuracy

- PG-HIVE achieves the **highest** F1 across datasets.
- **Robust** to noise.
- Works well even with **missing labels**.

Efficiency

- **Faster** than baseline methods.
- **Stable** runtime across noise levels.

Explore our results and test PG-HIVE with your own datasets!

Github

Contact us:

sophisid@csd.uoc.gr
kondylak@ics.forth.gr

[1]Sideri, S., Troullinou, G., Ymeralli, E., Efthymiou, V., Plexousakis, D., and Kondylakis, H.. PG-HIVE: Hybrid Incremental Schema Discovery for Property Graphs. EDBT 2026
[2]Sideri, S., Troullinou, G., Ymeralli, E., Efthymiou, V., Plexousakis, D., and Kondylakis, H.. 2025. PG-SB: Benchmarking Schema Discovery for Property Graphs doi:10.5281/zenodo.17801335
[3]Sideri, S., Efthymiou, V., Plexousakis, D., and Kondylakis, H.. 2025. PG2RDF: Schema-Guided Transformation of Property Graphs to RDF. In IJCKG 2025
[4]Kondylakis, H., Efthymiou, V., Troullinou, G., Ymeralli, E. ,and Plexousakis, D.. 2024. Property Graphs at Scale: A Roadmap and Vision for the Future (Short Paper)
[5]Kondylakis, H., Dumbrava, S., Lissandrini, M., Yakovets, N., Bonifati, A., Efthymiou, V., Fletcher, G., Plexousakis, D., Tommasini, R., Troullinou, G., and Ymeralli, E., 2025. Property Graph Standards: State of the Art & Open Challenges. PVLDB 18, 12 (2025)
[6]Lbath, H., Bonifati, A., Harmer, R.: Schema inference for property graphs. EDBT 2021
[7]Bonifati, A., Dumbrava, S., Mir, N.: Hierarchical clustering for property graph schema discovery. EDBT 2022
[8]Bonifati, A., Dumbrava, S., Martinez, E., Ghasemi, F., Jaffré, M., Luton, P., Pickles, T.: Discopg: Property graph schema discovery and exploration. Proc. VLDB 2022