Property Graph Encoding for Large Language Model

Advisors

- Angela Bonifati (Lyon 1 Univ., angela.bonifati@univ-lyon1.fr)
- Andrea Mauri (Lyon 1 Univ., andrea.mauri@univ-lyon1.fr)

Context

Graphs serve as unifying abstractions for encoding, inspecting, and updating interconnected data, with property graphs being the most expressive models to represent complex semantics and topologies. Property graphs are established as expressive models combining key-value pairs as properties with multiple labels on nodes and edges. Understanding graphs is crucial for making inferences about the connections between entities within a complex system and uncovering hidden patterns and trends.

LLMs have been applied in several domains, and recently also for reasoning over graphs [1,2], specifically on simple **labeled graphs** and considering a limited number of computational task types, such as verifying the existence of nodes, edges, cycles, and computing the degrees of a node.

However, their ability to reason over more expressive graph models (e.g., property graphs) is still uncertain. In particular, we are interested in understanding how the encoding of the graph changes the capability of the LLM to answer **queries** on the graphs, where queries can be about schema discovery, analysis, or aggregation.

Tasks

- Getting familiar with the property graph model [4] and the Neo4j database and the different types of query.
- Getting familiar with the state of the art regarding structured data [3] and graph text encoding and LLM prompting [1,2]
- Devise a set of extensive experiments to evaluate how different property graph encoding influences the capability of the LLM to reason over the graph.
- (Advanced) Evaluate other types of expressive graph model, such as RDF [5]

Requirements

- Good programming skills (Python, C++, Rust)
- Familiarity with property graph
- Familiarity with LLMs API (e.g., langchain).

References

[1] Fatemi, B., Halcrow, J., & Perozzi, B. (2023). Talk like a graph: Encoding graphs for large language models. arXiv preprint arXiv:2310.04560.

[2] Zhang, Y., Wang, H., Feng, S., Tan, Z., Han, X., He, T., & Tsvetkov, Y. (2024). Can LLM Graph Reasoning Generalize beyond Pattern Memorization?. arXiv preprint arXiv:2406.15992.

[3] Jiang, J., Zhou, K., Dong, Z., Ye, K., Zhao, W. X., & Wen, J. R. (2023, December). StructGPT: A General Framework for Large Language Model to Reason over Structured Data. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 9237-9251).

[4] Bonifati, A., Fletcher, G., Voigt, H., & Yakovets, N. (2022). Querying graphs. Springer Nature.

[5] https://www.w3.org/RDF/