Consistent-Preserving Interventions for Graph Queries

Supervisor : Angela Bonifati
Contact : angela.bonifati@univ-lyon1.fr
Laboratory : LIRIS UMR CNRS 5205
Master 2 : Master DISS (Data and Intelligence for Smart Systems)
Application deadline : January 31, 2027
Starting date : September 1, 2027

Abstract : (200 mots)
Property graphs model complex domains as vertices, edges, and properties, queried through pattern-matching languages such as Cypher. In settings like fraud analysis and knowledge graphs, understanding the sensitivity of query results to graph modifications is critical. Query causality formalizes this need: resilience measures the minimum set of edge deletions that eliminates all matches of a query, while causal responsibility quantifies how strongly an individual edge contributes to a result.
In practice, however, not every deletion is permissible. Graph databases enforce integrity constraints, in particular PG-key constraints, that must remain satisfied after any intervention. A naïve strategy repairs constraint violations after the fact, but this incurs substantial overhead and conflates two distinct concerns. This internship investigates a consistency-preserving approach that finds interventions satisfying minimality and constraint compliance simultaneously, treating them as a unified multi-objective optimization problem.
The work plan proceeds in four phases: formalizing constraint-aware resilience and causal responsibility and analyzing their computational complexity; implementing reproducible baselines from the state of the art; designing unified algorithms that enforce consistency during query evaluation; and evaluating effectiveness, efficiency, and scalability on synthetic and real-world graph datasets. The project is supervised by Prof. Angela Bonifati at LIRIS (Lyon 1), recipient of the ERC Advanced Grant Go-Y (https://projet.liris.cnrs.fr/erc-go-y/).

Project description :

Context and motivation

Property graphs represent entities as vertices, relationships as edges, and domain information as labels and properties. Languages such as Cypher retrieve answers by matching graph patterns. In fraud analysis, knowledge graphs, and social networks, users may also need to know how sensitive a result is to changes in the graph and which graph objects are critical to it. Query causality provides a principled framework for these questions. Resilience measures the smallest intervention

that invalidates a query result; causal responsibility measures how strongly an individual graph object contributes.

However, in real-world settings, not all interventions are permissible because of security, privacy, or accessibility restrictions. In graph databases, interventions must also respect user-defined integrity constraints to ensure consistent causal analysis. A naïve approach is to repair the database whenever an intervention introduces a constraint violation or inconsistency, but such repairs can impose substantial computational overhead. We therefore aim to develop a consistency-preserving approach that identifies only interventions guaranteed not to violate the specified constraints.

Problem and internship objective

The initial scope focuses on edge deletions and fixed conjunctive pattern queries. Given a graph G and a pattern query Q, each query match involves a set of edges. Resilience seeks a minimum-cardinality edge set S that intersects every match; deleting S therefore eliminates all matches of Q. For a given edge e, causal responsibility seeks a minimum-cardinality contingency set Γ such that Q still has a match after deleting Γ, but has no matches after additionally deleting e. If ($|\Gamma|=k$), the responsibility of e is $\frac{1}{1+k}$.

The identified intervention sets must also preserve database consistency. In particular, given user-defined PG-key constraints, an intervention must not introduce any constraint violations—that is, the corresponding violation-detection queries must return no matches. The minimality of an intervention and its consistency with the specified constraints must therefore be considered simultaneously.

This internship will investigate advanced methods for computing resilience and causal responsibility under consistency constraints. The goal is to identify interventions that preserve consistency while remaining as small as possible. The main challenge lies in solving this multi-objective optimization problem efficiently. Extensive experiments will evaluate the effectiveness, efficiency, and scalability of the proposed algorithms.



Figure: Ensure consistent-preserving interventions without additional repair.

Work plan

- **Foundations and formalization.** Review resilience, causal responsibility, and PG-key constraints; formally define the constraint-aware optimization problems and analyze their computational complexity.
- **Reference implementation.** Implement representative state-of-the-art methods and establish reproducible baselines for comparison.
- **Algorithmic design.** Develop unified, constraint-aware algorithms that compute interventions during query evaluation while jointly optimizing consistency and minimality.
- **Experimental evaluation.** Evaluate effectiveness, efficiency, and scalability on synthetic and real-world graph datasets, and compare the proposed methods with existing baselines.

Selected references

R. Angles, A. Bonifati, S. Dumbrava, G. Fletcher, K. W. Hare, J. Hidders, V. E. Lee, B. Li, L. Libkin, W. Martens, F. Murlak, J. Perryman, O. Savković, M. Schmidt, J. Sequeda, S. Staworko, and D. Tomaszuk. PG-Keys: Keys for Property Graphs. In Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21).

A. Meliou, W. Gatterbauer, K. F. Moore, and D. Suciu. The Complexity of Causality and Responsibility for Query Answers and Non-Answers. PVLDB 4(1), 2010.

C. Freire, W. Gatterbauer, N. Immerman, and A. Meliou. The Complexity of Resilience and Responsibility for Self-Join-Free Conjunctive Queries. PVLDB 9(3), 2015.

N. Makhija and W. Gatterbauer. A Unified Approach for Resilience and Causal Responsibility with Integer Linear Programming. ACM Transactions on Database Systems, 2023.

Neo4j. Cypher Manual: Patterns. https://neo4j.com/docs/cypher-manual/current/patterns/

Convertir le document au format pdf pour la soumission.