



Zürich, September 25, 2024

BSc Thesis

Topic: Implementing and Evaluating an Interval-Based Index for Finding Frontiers in Directed Acyclic Graphs

Tasks:

1. Literature Review:

- Study the relevant literature [1] to understand how the Interval-Based Index is built. This index will be used for solving the following frontier query problem: Let $G = (V, E)$ be a DAG and S be a subset of V . We are interested in all the vertices $s \in S$, for which there is no path in G from any other vertex in S to s .

2. Implement a brute-force BFS algorithm as a baseline

- In order to be able to compare the performance and to double-check the correctness of the frontier query that uses the index from [1], a naive version of the frontier query algorithm with breadth-first search needs to be implemented first.
- This also helps in setting up the infrastructure for benchmarking and testing the actual algorithm.

3. Preliminary evaluation:

- Set up an environment to evaluate the algorithms.
- This includes
 - the preparation of one or more datasets for processing
 - integrating the existing algorithms into the framework

4. Implement the interval-based data structure from [1]

- The next step is the implementation of the more efficient algorithm (from [1]).
- Depending on the outcome of a preliminary evaluation (mainly checking correct-

ness), some modifications of the algorithms may have to be introduced.

5. Full evaluation:

- The efficient algorithm is tested and evaluated using the same datasets as used in the preliminary evaluation. This time, the run time performance of the algorithm is also considered and compared to the brute-force baseline algorithm and the algorithms currently developed by Mateo.

6. Extending the algorithm (optional):

- Depending on how difficult the first tasks turn out to be and what insights are gained during the evaluation, the algorithm could be extended, for example by partitioning the DAG and building local indexes for the partitions.

7. Deliverables

- Write up your work and findings in a BSc thesis.
- Present your work in a talk at the database research group (optional).

Supervisor: Mateo Dujić, Sven Helmer

Start date: 01.10.2024

End date: 01.04.2025

References

- [1] R. Agrawal, A. Borgida, and H. V. Jagadish. Efficient management of transitive relationships in large data and knowledge bases. In *Proceedings of the 1989 ACM SIGMOD International Conference on Management of Data*, pages 253–262, Portland, Oregon, USA, 1989.

Department of Informatics, University of Zurich



Prof. Dr. Michael Böhlen

Professor