University of Zurich[UZH]

Department of Informatics

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Prof. Dr. Sven Helmer
Professor
Phone +41 44 635 67 28
Fax +41 44 635 68 09
helmer@ifi.uzh.ch

Zürich, 3. März 2026

MSc Thesis (30 points)
Topic: Optimizing Skip Table Building

Background and Motivation

Directed Acyclic Graphs (DAGs) are widely used to model dependencies in various domains. In this context, the Partial Sources Query problem involves finding all nodes reachable from a given subset of source nodes. The Layered Partial Sources with Skips (LPSS) algorithm optimizes such queries by utilizing precomputed skip tables based on the global depth (layering) of the DAG. The global depth of a node u is defined as the length of the longest path from any source node in the graph to u. While LPSS significantly improves query performance by utilizing these global skip pointers to bypass redundant traversals, building skip tables based purely on global depths can lead to a massive number of skip pointers. This results in a substantial memory footprint and slow index construction times.

Problem Statement

Computing skip tables based on global depths dictates that the index stores skip pointers covering long edges, connecting nodes with depth difference $>> 1$, at the same time with distance of only 1. This global view does not exploit local structural properties, leading to high redundancy. The challenge lies in reducing the number of skip pointers computed and stored, without significantly sacrificing the query performance of the LPSS algorithm.

Proposed Approach

The overarching goal of this thesis is to significantly optimize the process of building skip tables for the LPSS algorithm. To achieve a reduction in skip pointers and an acceleration of the index construction, we propose exploring the following main ideas:

1. **Local vs. Global Depths:** The first idea leverages the concept of *local depth*. We define the local depth of a node u with respect to a node v as the length of the longest path

from v to u. The goal is to reduce the number of skip pointers by computing the skip tables based solely on these local depths instead of global depths. Because the standard LPSS relies on global depths during query execution, this approach requires dynamically transforming local depths to global depths efficiently on-the-fly.

2. **Path-Based Skip Pruning:** The second idea introduces the concept of structural paths to identify redundant skip pointers. Specifically, if a node skips to an ancestor that resides on the same path, storing this skip pointer might be unnecessary since the path already provides the reachability structure. Pruning these redundant within-path skips could further drastically reduce the size of the skip tables.
3. **Further Optimizations:** In addition to the above, the project will leave room to explore and incorporate other topological heuristics and algorithmic refinements aimed at minimizing the index size without sacrificing the query latency of LPSS.

Primary Goals and Tasks

The project will focus on the following core objectives:

1. **Formalization:** Mathematically define local depth metrics, path-based pruning rules, and any additional optimization strategies for skip table construction.
2. **Algorithm Design:** Develop algorithmic frameworks for the optimized skip table construction (e.g., local-depth logic, path tracking) and formulate the necessary dynamic adaptation steps used during LPSS query execution.
3. **Implementation:** Implement the proposed construction methods and query transformation logic within the existing LPSS codebase.
4. **Evaluation:** Empirically evaluate the optimized approaches against the baseline global-depth LPSS algorithm, thoroughly analyzing memory footprint, index building time, and query performance across various real-world datasets.

Project Schedule (22 Weeks)

The thesis is planned to be completed over a 22-week period with the following tentative schedule:

- **Weeks 1–3:** Literature review, familiarization with the supervisor's previous work, and understanding the existing LPSS algorithm.
- **Weeks 4–7:** Formalization of local depth metrics and derivation of path-based skip pruning rules.
- **Weeks 8–11:** Algorithm design for optimized skip table construction.
- **Weeks 12–16:** Implementation of the new indexing method and query logic.
- **Weeks 17–19:** Empirical evaluation, pipeline setup, and compilation of benchmarking results.
- **Weeks 20–22:** Finalizing the analysis, drafting the written thesis, and preparing the final presentation.

University of Zurich
Department of Informatics

Prof. Dr. Sven Helmer
Professor

Supervisor: Mateo Dujić (`dujic@ifi.uzh.ch`)

Start date: March 3rd, 2026

End date: September 2nd, 2026

References