Scalable Overlay Operations over DCEL Polygon Layers

Andres Oswaldo Calderon Romero

University of California, Riverside

November 2024

Introduction

- Spatial data structures, especially DCEL, are essential for various spatial applications.
- Doubly Connected Edge List (DCEL) helps manage topological information of edges, vertices, and faces.
- Overlay operations enable integration of different spatial data layers.
- This work addresses scalability challenges for overlay computations over DCEL layers.

Motivation and Problem Statement

- Large spatial datasets (e.g., US census tracts) make traditional DCEL overlay operations challenging.
- Need for scalable, distributed DCEL overlay solutions.
- Objectives:
  - Efficient partitioning and merging for scalable DCEL overlays.
  - Enable parallel overlay operations while maintaining data integrity.

Related Work

- Historical use of DCEL in various applications (e.g., computational geometry, 3D graphics).
- Existing sequential algorithms for overlay operations: LEDA, CGAL.
- Need for scalable, parallel approaches in spatial data processing.

DCEL Overview

- DCEL consists of vertices, edges (half-edges), and faces.
- Each half-edge has references to twin, next, and previous edges.
- Allows efficient topological and geometric queries.

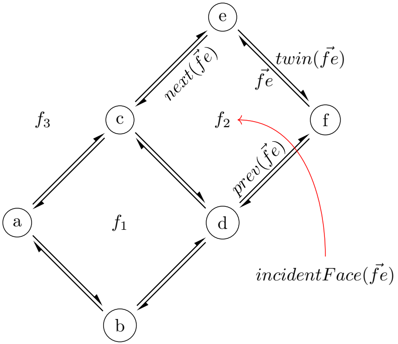

Figure: DCEL Structure

Scalable Overlay Construction

- Partition Strategy: Quadtree-based partitioning to manage large datasets.
- Each cell in the quadtree can be processed independently.
- Local results merged to form global overlay.

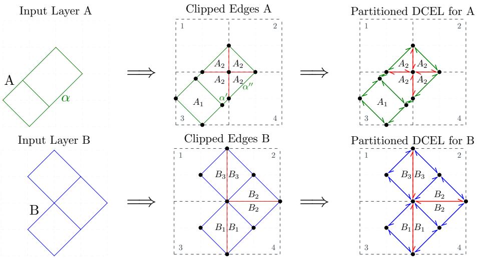

Figure: Partitioning Strategy for Overlay Computation

Handling Orphan Cells and Holes

- Problem: Holes or faces spanning multiple cells require special handling.
- Solution: Label propagation using quadtree structure.
- Recursive search for label consistency across partitions.

Optimization Strategies

- Optimization for Faces Spanning Multiple Cells
  - Intermediate reduce phase to handle large number of faces.
  - Spatial proximity property to minimize central node workload.
- Optimization for Unbalanced Layers
  - Focused scanning of dense layers to improve performance.

Experimental Evaluation

- Datasets: MainUS, GADM, CCT
- Evaluation metrics: Runtime, scalability, efficiency.
- Comparison with baseline approaches like CGAL.

Results and Observations

- Partitioning improves overlay operation efficiency.
- Optimizations reduce computation time, especially in unbalanced datasets.
- Scalable solution performs significantly better on large datasets than sequential methods.

Conclusions

- Developed a scalable overlay computation for DCEL layers.
- Approach supports distributed, efficient DCEL operations.
- Future work: Further improvements in handling extremely dense data.