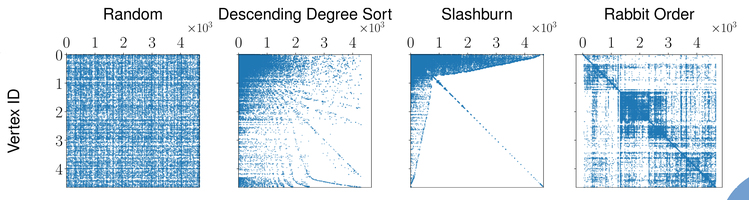
UBC Department of Computer Science

RHuBarb: Speeding up Edge-Centric Graph Processing Using Recursive Hilbert-Blocking

Alex Trostanovsky (atrostan@cs.ubc.ca), Margo Seltzer (mseltzer@cs.ubc.ca)

Problem

Vertex ID assignment drastically affects performance of graph analytics. **Given an input graph, how should we assign IDs to vertices?**



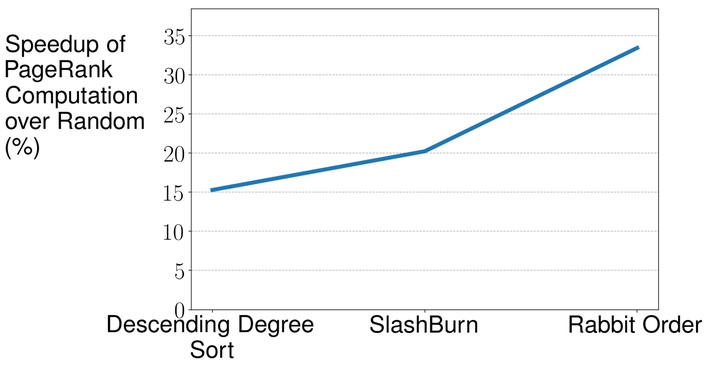


Figure 1. Different vertex orderings produce graph isomorphisms whose adjacency matrices contain distinct sparsity patterns.

Context

We focus on **Edge-Centric** graph analytics, where we compute properties of the graph by iterating over the edges.

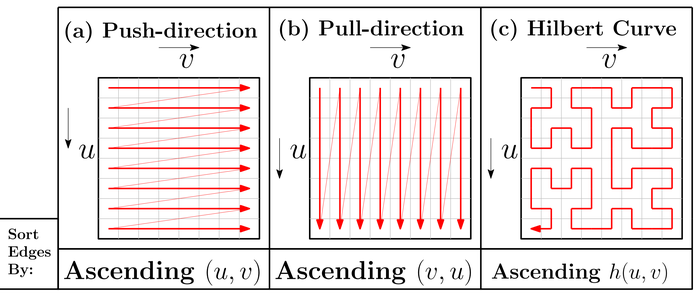
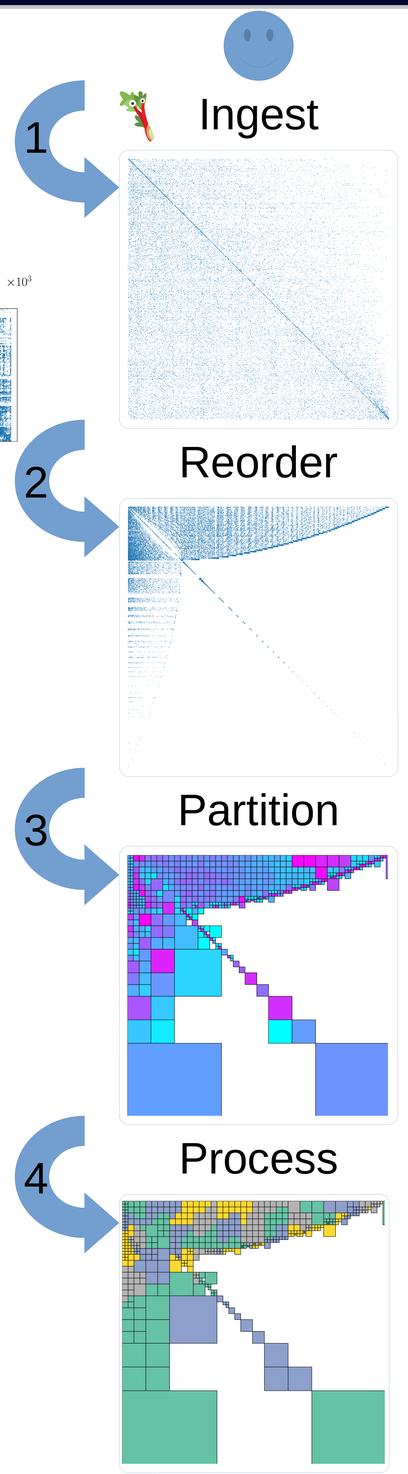


Figure 2. Common Edge Orderings



Our Approach - RHuBarb: Recursive Hilbert Blocking

1. Ingest and 2. Reorder

Given an input graph, **Rhubarb** ingests and (optionally) reorders its **vertices**. The graph is stored using a Compressed Sparse Row representation.

3. Partition

- Using Recursive Hilbert Blocking, we partition and reorder the **edges** of the graph for concurrent processing.
- **Rhubarb**'s divide-and-conquer approach ensures an even distribution of work among CPU cores.
- **Rhubarb** benefits from (but does not rely on) **compressed** graph representations.
- **Rhubarb** leverages the improved locality of the Hilbert Curve at the granularity of blocks and edges.

4. Process

- **Rhubarb** dynamically assigns cores to blocks, enabling the reuse of overlapping vertex data stored in shared caches.

Preliminary Results

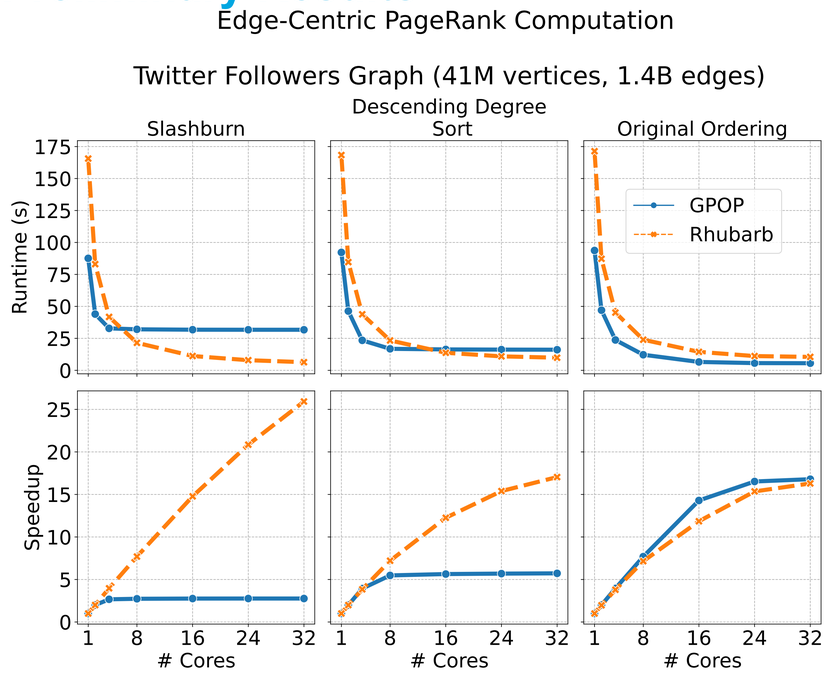


Figure 3. Runtime and Scaling of GPOP, Rhubarb. Predictably, GPOP is unable to scale for vertex orderings with concentrated edge densities, while Rhubarb benefits from such compressed representations, achieving near-linear scaling.

Contributions

- ***RHuBarb***: A Vertex-and-edge preprocessing pipeline that leverages:
  - Graph isomorphisms with concentrated edge densities,
  - **R**ecursive **H**ilbert **B**locking, a novel graph blocking approach, and
  - Concurrent, load-balanced edge-centric computations using a scalable parallelization of the Hilbert Curve.
- We identify a performance bottleneck on GPOP for graphs with concentrated edge densities, and show strong scaling on said graphs.

THE UNIVERSITY OF BRITISH COLUMBIA

Partners

NSERC CRSNG