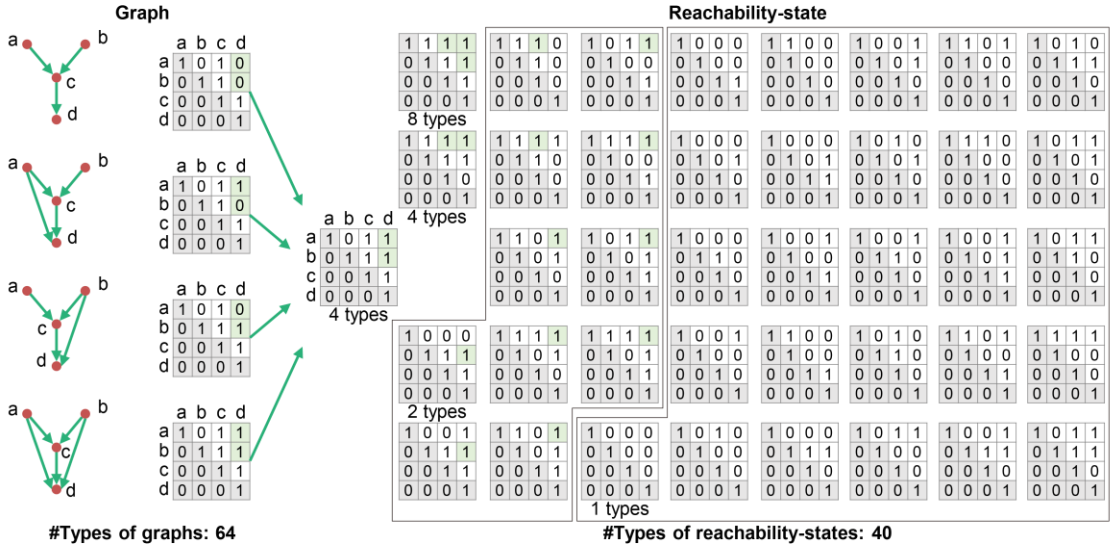
Reachability-state Encoding: Bridging tradeoffs between the Minimal Storage and the Fastest Query for Reachability Query

Huquan KANG, Xingli WANG, Zhouyang JIN, Yuang DING, Yiming LIU, Luoyi FU, Jiaxin DING, Xinbing WANG

Frontiers of Computer Science, DOI: 10.1007/s11704-025-50122-8

Problems & Ideas

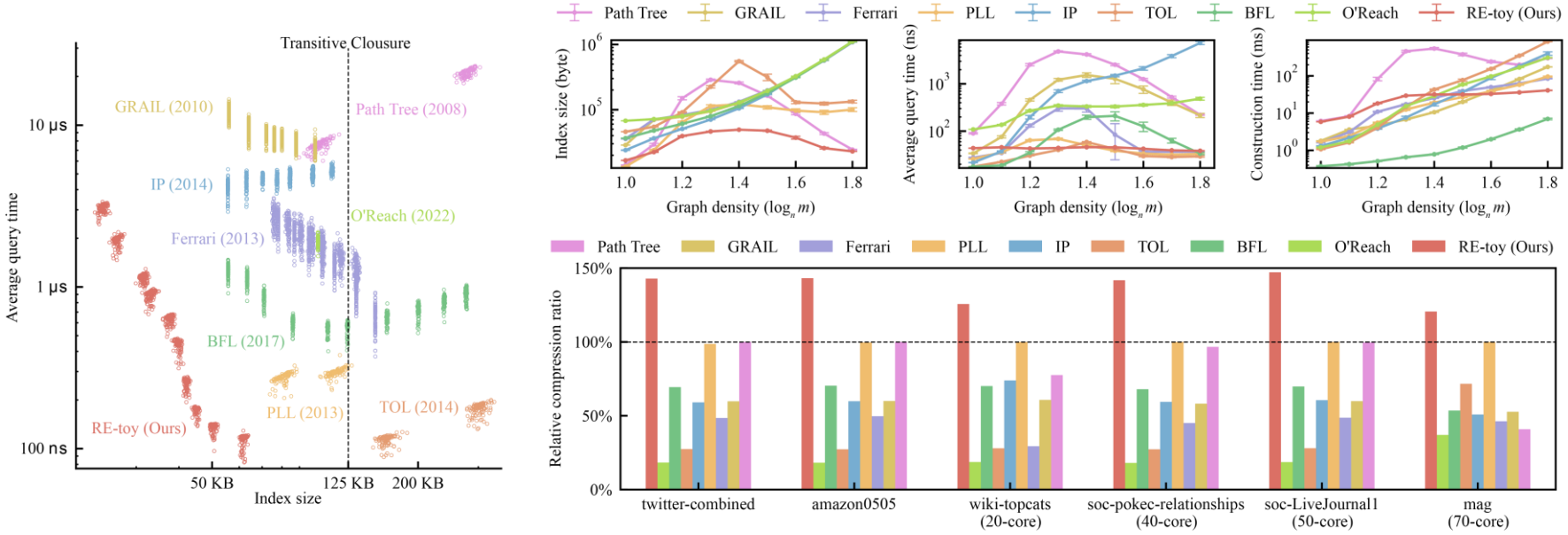
- Problems of conventional reachability query methods :
  - Traditional algorithms like DFS and transitive closure suffer from either slow query time or large storage overhead.
  - Existing indexing techniques often fix the tradeoff point, lacking flexible control over the balance between space and query efficiency.
- Ideas: Reachability-state Encoding (RE) rethinks graph indexing by compressing the transitive closure into a compact, entropy-guided format, enabling fine-grained control over storage-query tradeoffs.



Reachability-states on 4-node directed acyclic graphs. The 0/1 in the matrix of the graph indicates the existence of an edge, and 0/1 in the matrix of reachability-state indicates the existence of a path. The values in green cells do not affect the reachability-state of the graph, so a reachability-state may correspond to multiple types of graph adjacency matrices.

Main Contributions

- Contributions:
  - Propose the Reachability-state Encoding (RE) framework enabling smooth tradeoffs between space and time
  - Introduce reachability-state entropy to quantify the compressibility of reachability information
  - Develop the RE-toy prototype, demonstrating superior performance across diverse datasets



Efficiency and Robustness of RE-toy. Left: RE-toy achieves flexible tradeoffs between index size and query time, consistently approaching the optimal boundary. Right up: RE-toy maintains stable performance across varying graph densities, with low growth in space and time costs. Right down: On real-world datasets, RE-toy achieves over 30% average improvement in index size compared to the best existing algorithms.