

Memory-Constrained DiskANN

Yuhang Lou, Linyun Ma, Kun Luo, Yan Ruan, Huijia Wu,
Minhua Lu* and Rui Mao. Shenzhen University

Introduction

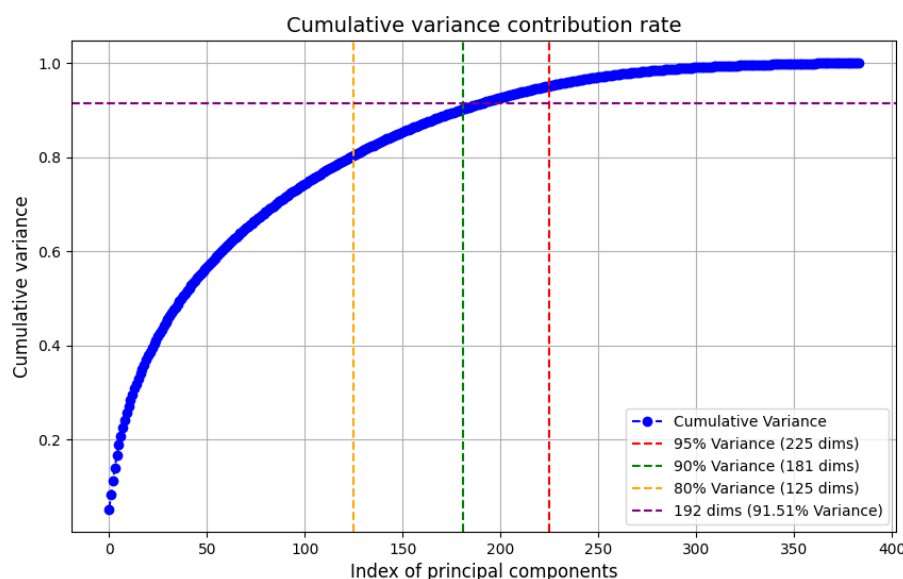
Using DiskANN to search over 23 million vectors within a 16 GB RAM limit, we apply PCA to reduce dimensionality, and introduce a selective second-assignment strategy. These optimizations cut memory and indexing costs while achieving over 80% recall—meeting the challenge’s performance goals.



Yuhang Lou

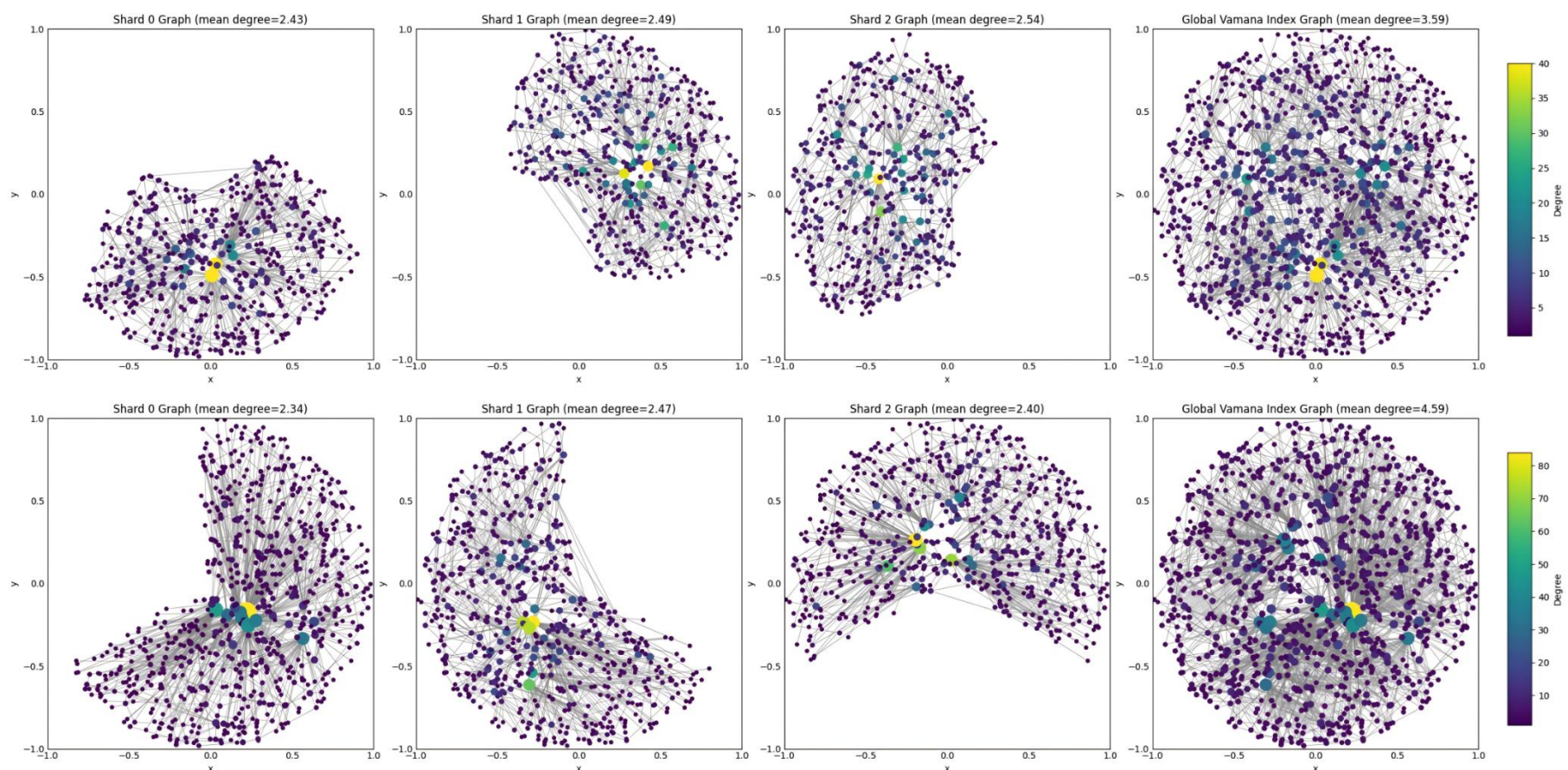
M.Sc. in CSSE, SZU

- Metric Space
- ANN Search
- Graph Generation



PCA-based dim reduction

- Sampled 10,000 base vectors to perform PCA
- From 384 to 192 dims preserves over 90% of total variance
- Lowers memory usage from over 16 GB to about 13.2 GB



Second-assignment strategy

- Assigned to its nearest, but do not all points need to the second for connection
- Reducing replication from $2\times$ to $1.5\times$ per point.
- Lowers memory usage, indexing time, and disk footprint, while maintaining high recall

