Approximate nearest neighbor search

- using Hierarchical Navigable Small World graphs

Group 20

Content

- Introduction
- Skip-Lists
- Hierarchical Navigable Small World graphs
- Similarity Search
- Algorithm Description
- Performance Evaluation
- Discussion

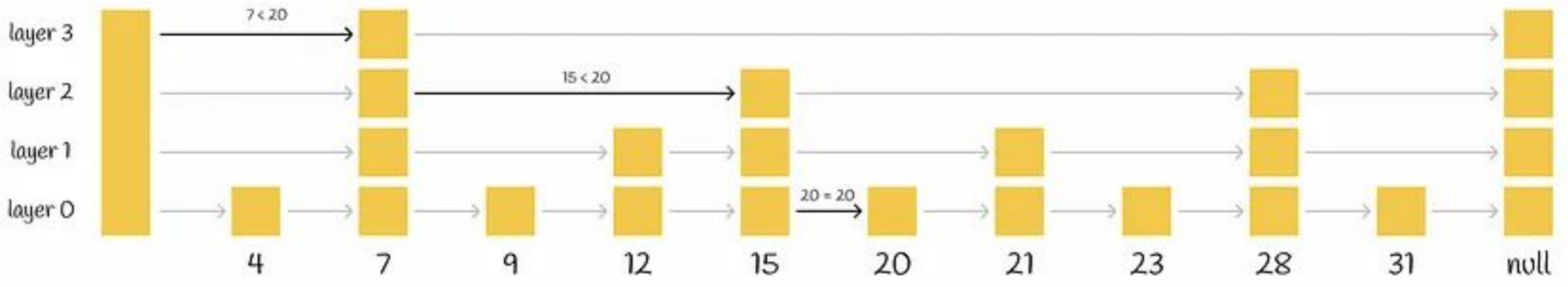
Introduction

- The constant growth of information

- Data Structure
  - Skip-List
  - Navigable Small World graphs (NSW)
  - Hierarchical Navigable Small World graphs (HNSW)

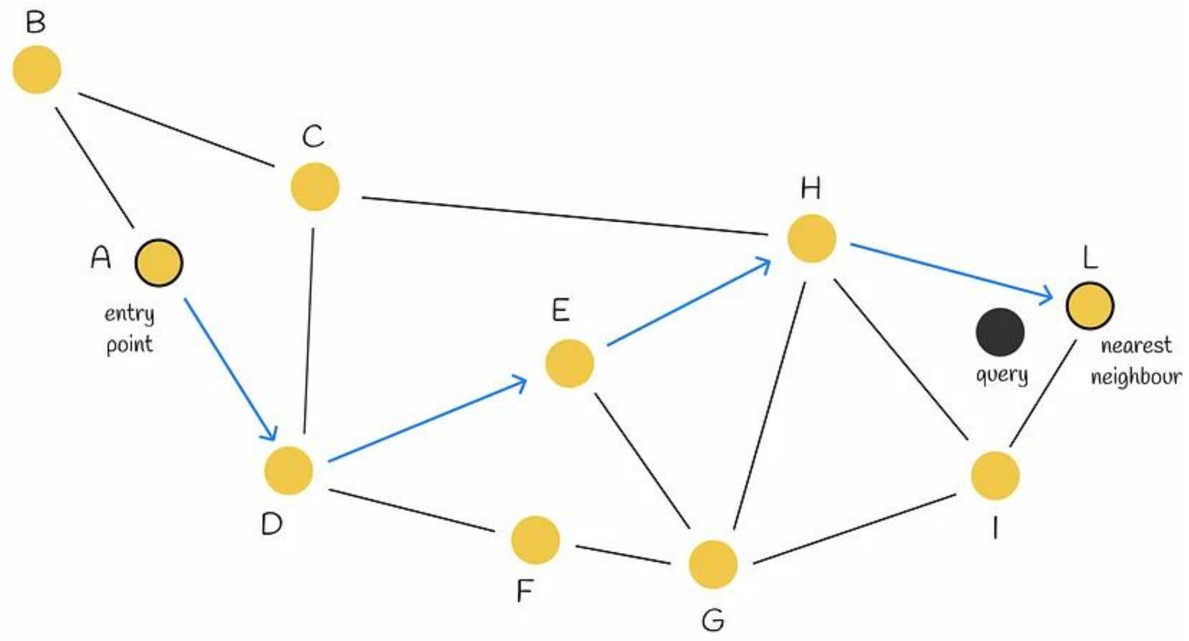
Skip-Lists

- Sorted linked lists
- O(log n)
- The search procedure



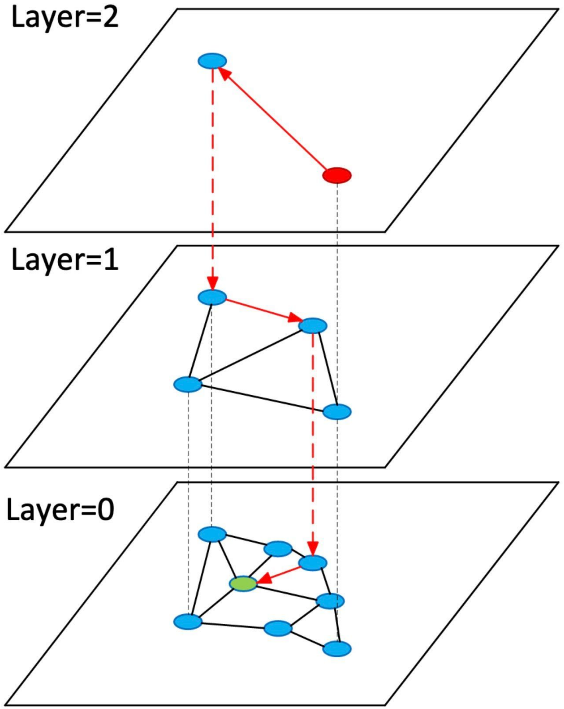
Navigable Small World graphs (NSW)

- Efficient pathfinding
- Small World
- Navigable



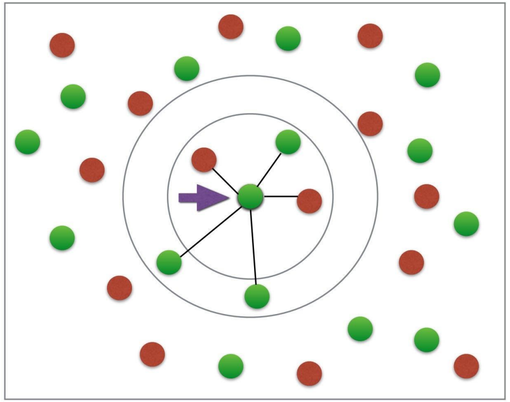
Hierarchical Navigable Small World graphs (HNSW)

- Multidimensional data
- Navigable Small World network
- Hierarchical approach
- Skip-Lists

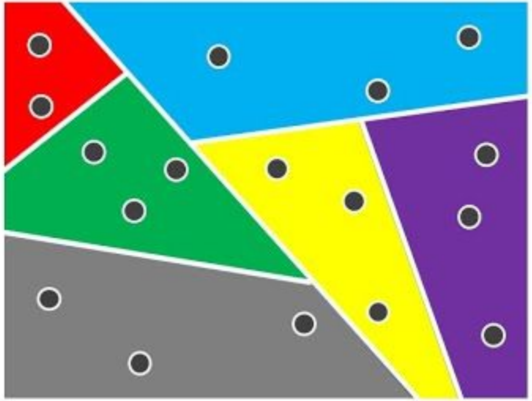


Similarity Search

- K-Nearest Neighbor Search (KNN)
  - Not feasible for large datasets
  - Accurate but heavy on resources



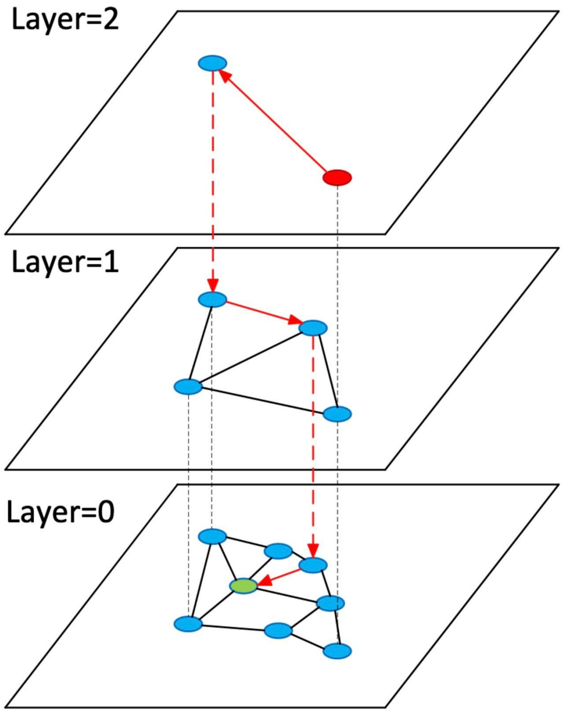
- Approximate Nearest Neighbor (ANN)
  - High dimensional datasets
  - Faster, but less accurate



Algorithm description

- Construction and search is similar
- Start at highest level (with the fewest nodes) L from entry node
- For each layer from top layer to layer 1:
  - Find the closest node to the node *q* being searched for
  - Continue searching at next level from the node found
- At level 0, find the k nearest nodes to *q* and return them

- For construction, uses probabilistic function to find which layer L to add new node to
- Adds new node to each layer L to 0

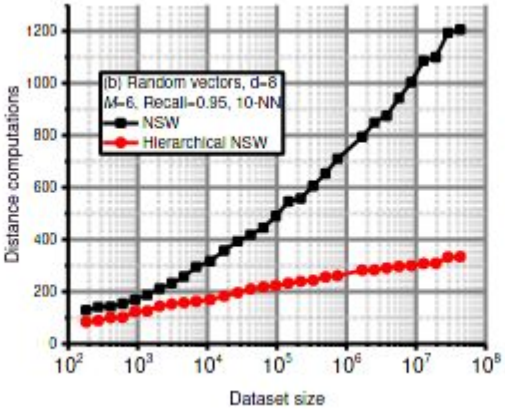


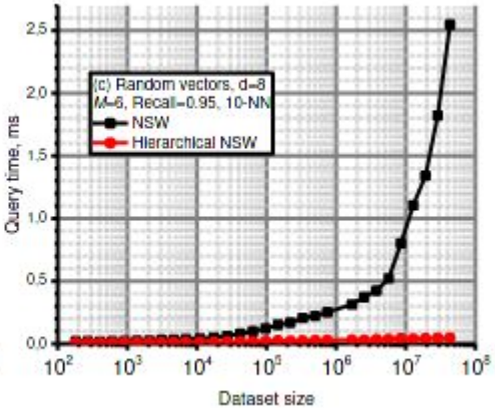
Complexity analysis

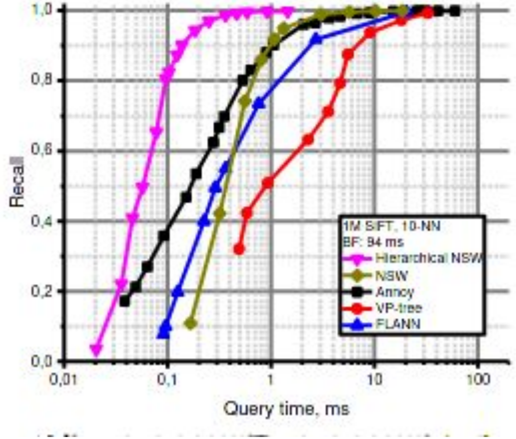
- Construction time scales as O(N log N)
- The average memory consumption per element is (Mmax0+mL·Mmax) ·bytes_per_link
- Skip list is O(log N)

Performance evaluation

- Construction time scales as O(N log N)
- The average memory consumption per element is (Mmax0+mL·Mmax)·bytes_per_link
- Skip list is O(log N)
- Tested against NSW and other state-of-the-art KNN algorithms
- Found to be faster and have higher recall than all the other algorithms







Discussion

- HNSW is very robust
- Shortcomming is it is not possible to do Distributed search, as was possible with NSW

References

- Y. A. Malkov and D. A. Yashunin, "Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 4, pp. 824-836, 1 April 2020, doi: 10.1109/TPAMI.2018.2889473.