Stochastic Traversal in Hierarchical Navigable Small World Graphs

Adi Pasic

December 14, 2022

1 Abstract

2 Introduction

High-dimensional search is concerned with finding the closest k matches to a query q in a database $\mathcal{D}$ of n elements. Applications include genetic sequence matching, video similarity checking, and related image search. Exact search requires comparison between the query and every object in the candidate set, which is computationally infeasible for large databases. Approximate search can be used in place of exact search if it estimates the k closest matches with relatively high accuracy.

Methods of search generally consist of partioning a set of candidates by proximity according to some distance metric. The data structures are constructed to exploit regularities in data proximity. A popular data structure is the Hierarchical Navigable Small World (HNSW), which has the property that greedy traversal finds a local minimum in logarithmic time.

3 Motivation and Background

Exhaustive search is equivalent to constructing a fully connected graph, in which case greedy search takes one hop but there are $n-1$ possible hops to evaluate. A key simplification is that for any database $\mathcal{D}$, there is a unique partitioning of $\mathbb{R}^d$ that segments regions according to which element they are nearest to, known as a Voronoi tessellation. If the Voronoi tessellation is known for a given graph, then the nearest-neighbor problem reduces to evaluating which Voronoi region the query belongs to. The best-known algorithm for constructing a Voronoi tessellation has $\mathcal{O}(n \log n)$ time complexity (citation needed, Fortune's Algorithm), so an approximation is necessary. The need for an approximation motivates study of the Delaunay triangulation, a graph that connects nodes to adjacent Voronoi regions. Greedy traversal over the Delaunay triangulation always yields the nearest neighbor to a query (citation needed).

Compared to exhaustive search, the Delaunay triangulation has much smaller average node degree but does not guarantee efficient greedy routing. A desirable property for greedy routing is that it can be done in $\mathcal{O}(\log^r(n))$ time, or polylogarithmic. Networks with this property are known as *navigable* small world (NSW) graphs.

Greeding routing in NSW has been empirically observed to have two phases: a phase with large hops and a phase with short hops (citation needed). This phenomena is tied to the structure of the graph as dense clusters joined with a few long-range edges between them. It takes relatively few edge insertions to dramatically reduce the average length between nodes.

4 Algorithm

The algorithm induces navigable small world properties by enforcing power-law scaling on node degrees and exponentially decaying probabilities on hierarchy. In previous work on constructing NSW, the insertion order of the graph was constructed in a

4.1 Complexity

The HNSW algorithm

Algorithm 1 HNSW Search

Require: q, ℓ
  for $\ell > 0$ **do**
    $c \leftarrow \text{GREEDY}(q, \ell)$
    $\ell \leftarrow \ell - 1$
  end for
  return c

Algorithm 2 HNSW Construction

Require: q, m_ℓ, M, ef
  $\ell \leftarrow \text{Exp}(m_\ell)$
  for $\ell > 0$ **do**
  end for

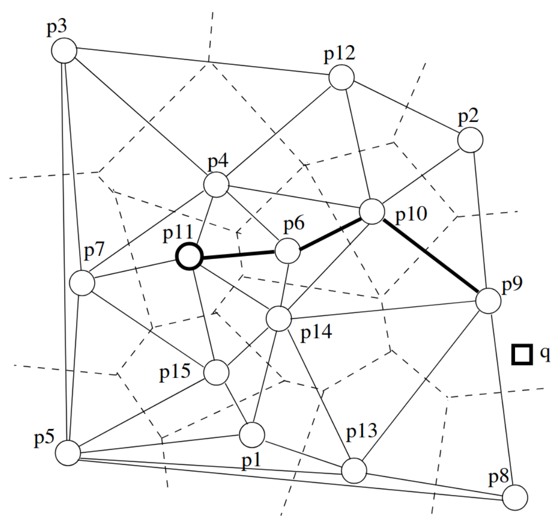

Figure 1: An illustration of the Voronoi tessellation and corresponding Delaunay triangulation on a set of 15 nodes. Greedy traversal on a Delaunay graph always finds the nearest neighbor to a query. Figure obtained from [Nav02]

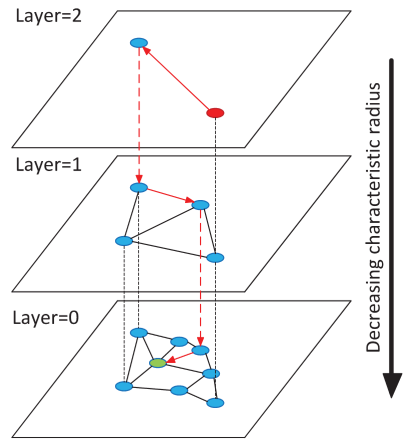

Figure 2: An illustration of traversal in HNSW. Search begins at the red node in the top layer. The trajectory of greedy traversal is indicated by red arrows. After a local minimum is reached, search continues in the next layer. Figure obtained from [MY18]

5 Algorithm

Let $\mathcal{D}$ be a database with n vectors $x_1, \ldots, x_n \in \mathbb{R}^d$.

6 Simulation Results

References

[Nav02] Gonzalo Navarro. "Searching in Metric Spaces by Spatial Approximation". In: *The VLDB Journal* 11.1 (2002), pp. 28–46.

[MY18] Yu A Malkov and Dmitry A Yashunin. "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs". In: *IEEE transactions on pattern analysis and machine intelligence* 42.4 (2018), pp. 824–836.