UNDERSTANDING HUBNESS FOR GRAPH-BASED NEAREST NEIGHBOUR SEARCH

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph-based approximate nearest neighbour search algorithms have gained significant attention recently due to their excellent empirical performance, though theoretical understanding of their behaviour remains limited. Empirical work has established the importance of 'hubs' during the greedy search. In this work, we offer some theoretical explanation for the 'hubness' phenomenon and its importance for nearest neighbour search. We develop a large-deviation framework for Gaussian data in the moderate high-dimensional regime $d = \Theta(\log(n))$. Through analysis of greedy nearest-neighbour search on data drawn from the Gaussian and uniform spherical distributions, we distinguish geometric hubs, characterized by high incoming degree, from navigation hubs, characterized by more frequent visitation during search. Together, these results connect graph geometry, hubness, and navigation guarantees, providing analytical tools for distribution-sensitive theories of graph-based nearest neighbour search.

1 INTRODUCTION

Approximate nearest neighbour search (ANNS) is a basic geometric operation on data. Given a finite dataset $\mathcal{X} = \{X_1, \ldots, X_n\} \subset \mathbb{R}^d$ and a query point $q \in \mathbb{R}^d$, the ANNS problem asks for a point in $\mathcal{X}$ that is close enough to q under a prescribed distance. This is an important subroutine used in information retrieval, recommendation systems, machine learning, data mining , and vector databases. Graph-based methods provide a strong empirical trade-off between search efficiency and accuracy, especially for large, high-dimensional datasets where high recall is required (Wang et al., 2021; Li et al., 2019). HNSW, NSG and DiskANN are representative methods which construct proximity graphs over the dataset and use query distance to guide the search (Malkov & Yashunin, 2018; Jayaram Subramanya et al., 2019; Fu et al., 2019). In graph-based methods, the search is implemented with a simple greedy search, in which at each step, all outoing neighbors of the current vertex are examined and the closest one to the query is chosen.

An important feature of such state-of-the-art graphs is *hubness*: some points appear abnormally often in the nearest-neighbour lists of other points. Previous work relates this imbalance to spatial centrality, where points lying closer to the center of the data tend to be closer to many other points and therefore appear more frequently as nearest neighbors (Radovanovic et al., 2010; Angiulli, 2018). Recent empirical work has connected hubness to the behavior of graph-based search. Munyampirwa et al. (2026) argue that frequently traversed hubs form a 'highway', which helps explain why flat proximity graphs can achieve performance comparable to HNSW. Xu et al. (2026) investigate how hubs and anti-hubs affect search efficiency, especially for outlier queries. These insights call for a deeper theoretical understanding of how the geometric advantage of receiving more incoming edges relates to search.

Existing theoretical analyses provide a natural starting point. Laarhoven (2018) study time–space trade-offs for ANNs on threshold graphs in high dimensions by considering the volumes of spherical caps and their intersections. Prokhorenkova & Shekhovtsov (2020) use similar techniques to analyze several practical search heuristics. Both works consider spherical uniform data, where the neighborhood geometry is identical at every vertex. On the other hand, although Gaussian norms concentrate around$\sqrt{d}$, this does not remove radial variation at the scale of the whole dataset. In this paper we show that the variation in the norm of the data points plays an important role during search.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We consider n independent points drawn from $\mathcal{N}(0, I_d)$ in the moderate high-dimensional regime $d/\log n \to w$ for a constant w. We use the large deviation principle (LDP) to characterize the geometric structure of the graph, and develop a framework for analyzing graph-based ANNS, which we apply to both Gaussian and spherical uniform data.

The main contributions of this work are as follows.

1. In the moderate-high dimensional Gaussian regime, there is a function $\rho_{\mathrm{crit}}(c, w) < 1$ such that for any $\rho_k \in (\rho_{\mathrm{crit}}(c, w), 1)$, $c > 1$ and w above an explicit constant w_0, greedy search on the directed k-NN graph with degree $k = n^{\rho_k + o(1)}$ encounters a c-approximate nearest neighbor within $O(1)$ steps with high probability. This is stated formally in Theorem 1, and we explore the properties of ρ_{crit} in section 4.1.
2. We identify a distinction between 'geometric hubs' (vertices with high in-degree) and 'navigable hubs' (high in-degree vertices which are visited by greedy search). Counter-intuitively, in Section 4.2 we show that the hubs with highest in-degree in the dataset are never visited by greedy search.
3. Our proof makes use of the large deviation principle and establishes a general technique for asymptotic analysis of greedy search on k-NN graphs which may be of independent interest.
4. We apply this framework to spherical uniform data (which does not have hubs) and derive the degree exponent $\rho_{\mathrm{uni}}(c, w)$ which is sufficient for c-ANN. In Section 4.3, we show that the Gaussian model requires a higher graph degree in low dimensions but becomes more efficient when w is sufficiently large. This provides further theoretical evidence that navigable hubs are key to accelerating graph-based nearest neighbour search in high dimensions.

2 Related Work

The theoretical work closest to ours focuses on greedy search on near-neighbor graphs for spherical uniform data. Laarhoven (2018) analyze an α-near-neighbor graph, where two spherical data points are connected when their inner product exceeds a fixed threshold. An improving neighbour must lie both in the local cap around the current vertex and in the cap containing points closer to the query. Their intersection therefore determines the probability of finding an improving neighbor. The analysis separates successive local improvements from a final jump to a planted near neighbor, and includes random restarts when deriving the time–space trade-offs. They consider both $n = 2^{o(d)}$ and $n = 2^{\Theta(d)}$, the latter corresponding to the moderate regime in our terminology. Prokhorenkova & Shekhovtsov (2020) focus mainly on the dense regime $d \ll \log n$ and analyze a single search tour. Beyond plain greedy search, they investigate how shortcut edges accelerate navigation and how maintaining several promising candidates through beam search improves accuracy. Shrivastava et al. (2023) further study approximate near-neighbor graphs, where some edges of the underlying threshold graph may be missing. Their analysis quantifies how retaining fewer neighbor edges affects the probability of successful search. Our analysis follows the same threshold-graph viewpoint, but the Gaussian neighborhood radius depends on the source norm and therefore requires a source-dependent method.

Hubness is commonly studied through the number of k-occurrences, which counts how often a point appears in the k-nearest-neighbor lists of other points. Radovanovic et al. (2010) show that, under commonly used distributional assumptions, this distribution becomes increasingly skewed as dimension grows, producing hubs with unusually many occurrences. Angiulli (2018) give a more quantitative description through expressions for distance distributions, nearest-neighbor distances, and the number of k-occurrences associated with a point. These results provide a geometric perspective on neighbor popularity and its relation to spatial centrality. Recent work examines how this popularity relates to graph search. Munyampirwa et al. (2026) compare HNSW with its flat base graph and find similar latency and recall on their high-dimensional benchmarks. They also study visited node patterns and connectivity, which provide evidence that frequently traversed vertices form a well-connected 'hub highway'. Xu et al. (2026) investigate the difficulties associated with overly concentrated connectivity and isolated anti-hubs. They relate these structures to slow queries targeting anti-hubs and propose hubness-aware adaptive pruning to improve graph topology. These studies motivate a

distinction between static incoming-edge popularity and the vertices preferred by query-dependent greedy search.

The relation between dimension and dataset size is also important when interpreting high-dimensional concentration. Beyer et al. (1999) give conditions under which the relative difference between nearest and farthest-neighbor distances vanishes as dimension grows. Giannella (2021) study Gaussian data with a growing dataset size and give sufficient conditions for both nearest-neighbor instability and its failure. In our regime, $n = \exp(d/w + o(d))$, so events with probabilities exponentially small in d can still appear many times in the dataset. This makes the large deviation principle(LDP) a natural tool for describing the relevant neighbourhood geometry (Dembo & Zeitouni, 2009).

3 PRELIMINARIES

3.1 PROBLEM SETTING

We consider a dataset $\mathcal{X} = \{X_1, \cdots, X_n\}$ of independent Gaussian points,

$$X_i \sim \mathcal{N}(0, I_d), \qquad \frac{d}{\log n} \longrightarrow w \in (0, \infty). \tag{1}$$

The parameter w is fixed as $n, d \to \infty$, and $\log(\cdot)$ is the natural logarithm throughout the paper. The query q is independent of $\mathcal{X}$ and satisfies $\|q\|^2/d \to 1$ in probability[1]. Given $c > 1$, the c-approximate nearest neighbour (c-ANN) problem asks for a point $X_j \in \mathcal{X}$ such that

$$\|X_j - q\| \leq c \min_{i \leq n} \|X_i - q\|. \tag{2}$$

We say that a statement holds with high probability if its probability tends to one as $n \to \infty$.

The directed k-nearest-neighbour graph on $\mathcal{X}$ connects each point to its k closest other data points. Following the degree parametrization of Laarhoven (2018) and Prokhorenkova & Shekhovtsov (2020), we parametrise the degree as $k = n^{\rho_s + o(1)}$, where $0 < \rho_s < 1$. We study the greedy search algorithm for c-ANN, in which at each iteration, the algorithm examines the outgoing neighbors of the current vertex v and moves to the one closest to q, provided that its distance is smaller than $\|v - q\|$. Otherwise, it stops.

3.2 LARGE DEVIATION PRINCIPLE AND SURROGATE MODEL

Rather than analyzing the k-NN graph directly, we will follow prior work (Laarhoven, 2018) and instead use a surrogate model in which any given point v is connected to every other point which lies within some specified threshold distance of v. In contrast with previous work, the threshold distance varies for different v. Specifically, it is a function of $\|v\|$.

Definition 1 (The τ-threshold graph)**.** *Let $\tau(\alpha)$ be the unique value such that*

$$\mathbb{P}\left[\frac{\|U - v\|^2}{d} \leq \tau(\alpha)\right] = \frac{k}{n-1},$$

when v is a fixed vector with $\|v\|^2 = d\alpha$, and $U \sim \mathcal{N}(0, I_d)$[2]. We define the τ-threshold graph on data set $\mathcal{X}$ to be a directed graph $G_\tau(\rho_s) = (V, E)$ with vertex set $V := (X_1, ..., X_n)$. E contains the directed edge $X_i \to X_j$ if and only if

$$\frac{\|X_i - X_j\|^2}{d} \leq \tau\left(\frac{\|X_i\|^2}{d}\right).$$

To establish suitable values of $\tau(\alpha)$, we need to know how likely an independent point is to lie within a certain distance of v. Since we would like k neighbours, the probability should be

$$\frac{k}{n-1} = \exp\left(-\frac{d(1-\rho_s)}{w} + o(d)\right), \tag{3}$$

[1]For example, consider $q \sim \mathcal{N}(0, I_d)$.

[2]Notice that by the rotational symmetry of $\mathcal{N}(0, I_d)$, the direction of v does not affect this probability

which is exponentially small in d. The large deviation principle (LDP) describes probabilities on this scale through a rate function (Dembo & Zeitouni, 2009). If a probability measure (also called **law**) μ satisfies an LDP with speed d and rate function Λ, then

$$-\inf_{\varrho\in F^\circ}\Lambda(\varrho)\le\liminf_{d\to\infty}\frac{1}{d}\log\mu(F)\le\limsup_{d\to\infty}\frac{1}{d}\log\mu(F)\le-\inf_{\varrho\in\bar{F}}\Lambda(\varrho).$$

where F° and $\bar{F}$ denote the interior and closure of F, respectively. If $\inf_{\varrho\in F^\circ}\Lambda(\varrho)=\inf_{\varrho\in\bar{F}}\Lambda(\varrho)$ we can obtain the limit and rewrite it as

$$\lim_{d\to\infty}\frac{1}{d}\log\mu(F)=-\inf_{\varrho\in F}\Lambda(\varrho)\iff\mu(F)=\exp\left(-d\inf_{\varrho\in F}\Lambda(\varrho)+o(d)\right).$$

Thus the rate of an event is obtained by minimising the rate function over that event. Note that a larger rate means a smaller probability. For n independent candidates, a region with probability $\exp(-dI+o(d))$ contains an expected

$$n\exp(-dI+o(d))=n^{1-wI+o(1)}$$

points. So, if $I<1/w$, the region contains a candidate with high probability and if $I>1/w$, it is empty with high probability. The boundary $I=1/w$ needs more than the leading probability exponent to determine whether a candidate occurs.

We now apply this calculation to Gaussian neighbourhoods. For a source v, write $\alpha=\|v\|^2/d$. Rotational symmetry makes the neighbourhood probability depend on α alone. For an independent point $U\sim\mathcal{N}(0,I_d)$, we establish in Theorem 6 that the normalised squared distance $\|U-v\|^2/d$ has rate function

$$I_\alpha(s)=\frac{m-1-\log m+\alpha(1-m)^2}{2},\qquad m=m_\alpha(s):=\frac{2s}{1+\sqrt{1+4\alpha s}},\quad s>0.\tag{4}$$

We set $I_\alpha(s)=+\infty$ for $s\le0$. In particular, for $0<s<1+\alpha$,

$$\mathbb{P}\left(\frac{\|U-v\|^2}{d}\le s\right)=\exp\left(-dI_\alpha(s)+o(d)\right).$$

On this interval, I_α is continuous and strictly decreasing from $+\infty$ to 0. We therefore define two unique distances $s_*(\alpha;\rho)$ and $s_{\mathrm{NN}}(\alpha)$ such that

$$I_\alpha\big(s_*(\alpha;\rho_s)\big)=\frac{1-\rho_s}{w},\qquad I_\alpha\big(s_{\mathrm{NN}}(\alpha)\big)=\frac{1}{w}.\tag{5}$$

The first distance gives a neighbourhood threshold distance which contains $n^{\rho_s+o(1)}$ points in expectation. So, to build a τ-threshold graph to represent the k-NN graph, we will set $\tau(\alpha)=s_*(\alpha;\rho)$. The second distance is the expected distance from a point to its nearest-neighbour. Both are normalised distances, so the corresponding Euclidean distances are $\sqrt{d\,s_*(\alpha;\rho_s)}$ and $\sqrt{d\,s_{\mathrm{NN}}(\alpha)}$.

A well known property of the Gaussian distribution in high dimensions is the concentration of the samples' norms (François et al., 2007). However, in our moderate regime with $d=\Theta(\log(n))$, the distribution of norms is important and we characterise its interval. The normalized squared norm $\|X_i\|^2/d$ can be seen as a special case of (4) and has rate

$$I_0(\alpha)=(\alpha-1-\log(\alpha))/2.$$

Let $\alpha_{\min}<1<\alpha_{\max}$ be the two solutions of $I_0(\alpha)=1/w$. We have $I_0(\alpha)\in[0,1/w]$, $\forall\alpha\in[\alpha_{\min},\,\alpha_{\max}]$. We show in Lemma 7 that for any constant $h>0$,

$$\mathbb{P}\left(\frac{\|X_i\|^2}{d}\in[\alpha_{\min}-h,\alpha_{\max}+h]\text{ for every }i\le n\right)\longrightarrow1.$$

We will use $[\alpha_{\min},\alpha_{\max}]$ as the norm interval which holds in the limit. From the analysis above, the nearest-neighbour distance satisfies

$$\min_{i\le n}\frac{\|X_i-q\|^2}{d}\to s_{\mathrm{NN}}(1).$$

We therefore define the c-ANN target distance to be $T:=c^2\cdot s_{\mathrm{NN}}(1)$. For a current vertex v, write $r:=\frac{\|v-q\|^2}{d}$. Thus to demonstrate that greedy search is successful, we must show that we make progress from every feasible non-target state

$$\alpha\in[\alpha_{\min},\alpha_{\max}],\qquad r>T,$$

until the target region $r\le T$ is reached.

3.3 Geometry of One Greedy Step

An improving neighbour must meet three requirements. It must occur in the dataset, lie within the neighbourhood of the current vertex v, and be closer to the query q. We now formalise these requirements using the position of a candidate relative to the v and q.

We will assume that the query norm is $\beta := \|q\|^2/d = 1$, although the results will extend to β sufficiently close to one, as shown in Lemma 19. Write $\theta := \langle v, q\rangle/d$. We will assume, without loss of generality, that

$$v = \sqrt{\alpha d}e_1\,, \quad q = \sqrt{\frac{d}{\alpha}}\theta e_1 + \sqrt{d\left(1-\frac{\theta^2}{\alpha}\right)}e_2,$$

where $e_1, \dots e_d$ are the standard basis vectors. Geometric feasibility gives $\alpha > 0$ and $\theta^2 \le \alpha$. Then we will define $r = \alpha + 1 - 2\theta$ to be the current distance from the query.

For an independent point $U \sim \mathcal{N}(0, I_d)$ it is natural to consider the projection of U onto the first two basis vectors, and the remaining dimensions:

$$A_d = \frac{\langle U, e_1\rangle}{\sqrt{d}}, \quad B_d = \frac{\langle U, e_2\rangle}{\sqrt{d}}, \quad M_d = \frac{1}{d}\sum_{j=3}^{d}\langle U, e_j\rangle^2.$$

In Theorem 5 we establish that the probabilistic density of (A_d, B_d, M_d) satisfies an LDP on $\mathbb{R}^3$ with speed d and rate function

$$J(a,b,m) := \begin{cases} \frac{1}{2}(a^2+b^2+m-1-\log m), & m > 0 \\ +\infty, & m \le 0 \end{cases} \tag{6}$$

For a candidate $x = (a, b, m)$, define the local term and greedy term by

$$g_l(x;\alpha) := (a-\sqrt{\alpha})^2 + b^2 + m, \quad g_G(x;\alpha,\theta) := \left(a - \frac{\theta}{\sqrt{\alpha}}\right)^2 + \left(b - \sqrt{1-\frac{\theta^2}{\alpha}}\right)^2 + m. \tag{7}$$

Viewing the candidate as (A_d, B_d, M_d), these are exactly $\frac{\|U-v\|^2}{d}$ and $\frac{\|U-q\|^2}{d}$. Thus, for each feasible non-target state, we consider the following continuous optimization problem:

$$\min_{x\in\mathbb{R}^3} g_G(x;\alpha,\theta) \;\text{ subject to }\; J(x) \le \frac{1}{w},\; g_l(x;\alpha) \le s_*(\alpha;\rho_s). \tag{8}$$

The constraint on J describes the range of available candidates. The local constraint on g_l restricts them to the neighbourhood of v and the objective selects the candidate closest to the query. To obtain an improving neighbour with high probability, we will construct a candidate x satisfying the strict inequalities

$$J(x) < \frac{1}{w}, \qquad g_l(x;\alpha) < s_*(\alpha;\rho_s), \qquad g_G(x;\alpha,\theta) < r.$$

These inequalities give a small region of improving candidates with enough probability mass to be represented in the dataset.

4 Main Results

In this section, we state and discuss our main theoretical results. We first give a condition on the degree required for greedy search on the k-NN graph on Gaussian data. Then we define and study geometric and navigational hubs in Gaussian data and demonstrate that while search on the k-NN graph does make use of hubs, it does not visit the vertices with highest in-degree. Finally, we compare the degree requirement for Gaussian data with that of spherical uniform data and establish that the additional hub structure in Gaussian data accelerates nearest neighbour search when the dimension is sufficiently large. Throughout this section we present the key insights and defer the detailed proofs to the appendix.

4.1 GREEDY SEARCH ON GAUSSIAN DATA

Let w_0 be the unique positive solution of $\alpha_{\min} + s_{\mathrm{NN}}(1) = 1$ and let $\rho_{crit}(c, w)$ be a function which we define later in equation (11) .

Theorem 1 (Gaussian Navigation)**.** *Fix $w \geq w_0$, $c > 1$ and $\rho_{crit}(c, w) < \rho_s < 1$. With high probability, greedy search on $G_\tau(\rho_s)$ from any starting vertex finds a c-approximate nearest neighbour of q within $O(1)$ steps.*

The formal proof can be found in Appendix C, Theorem 10. The remainder of this section shows the key ideas used in the proof.

Analysing the original optimisation problem (8) is difficult. At a high level, our goal is to minimise the value of ρ_s that allows the greedy search to make progress at every step, but notice that ρ_s appears inside a constraint and so changing ρ_s changes the feasible region. As such, we shift our perspective and instead define a problem to directly minimize ρ_s (or equivalently, to minimize the local threshold distance) while allowing us to make progress towards the query at every step. For a geometrically feasible state with $r > s_{\mathrm{NN}}(1)$, define

$$s_{\mathrm{crit}}(\alpha, r) := \min_{x \in \mathbb{R}^3} \left\{ g_l(x; \alpha) : J(x) \leq \frac{1}{w}\,, \quad g_G(x; \alpha, \theta) \leq r \right\}, \quad \theta = \frac{\alpha + 1 - r}{2}\,. \tag{9}$$

This is the smallest local threshold distance that guarantees a greedy search candidate exists (since $J(x) \leq 1/w$) which is at least as close to q as the current vertex. For now, we allow $r = g_G(x; \alpha, \theta)$, but later the strict inequality $\rho_s > \rho_{\mathrm{crit}}$ establishes that we make constant progress at every step. This new objective (9) is solvable, and the details are in appendix B. Recalling that there is a correspondence between the degree exponent ρ, and the threshold distance s, we obtain

$$\rho_{\mathrm{crit}}(c, w) := \sup_{\substack{\alpha \in [\alpha_{\min}, \alpha_{\max}] \\ r > T}} \left[1 - w I_\alpha \left(s_{\mathrm{crit}}(\alpha, r)\right)\right] \tag{10}$$

For $w \geq w_0$, the inside of the supremum is non-increasing in both α and r. This simplifies the definition to be

$$\rho_{\mathrm{crit}}(c, w) = 1 - w I_{\alpha_{\min}}(s_{\mathrm{crit}}(\alpha_{\min}, T)) \tag{11}$$

Remark 1. *The fact that the critical degree exponent is non-increasing in α and r gives some insight into the navigability of the k-NN graph. Firstly, when the current state is close to the query (r is small) we need a higher degree in order to make progress and secondly, when the norm of the current vertex is small (α is small) we need a higher degree to make progress. This suggests that lower-norm regions are not necessarily easier to navigate, despite their larger incoming degree. We return to this point in Section 4.2.*

Remark 2. *Although ρ_{crit} is well defined in terms of the LDP rate function, it is difficult to analyse symbolically. In Section 5, we include some numerical results to show the shape of ρ_{crit} as a function of c and w.*

The strict gap $\rho_s > \rho_{\mathrm{crit}}(c, w)$ in Theorem 1 is important. It gives a strict inequality between the actual neighbourhood radius $s_*(\alpha; \rho_s)$ and the critical radius $s_{\mathrm{crit}}(\alpha, r)$ at every non-target state. By exploiting this strict inequality, a union bound over all current vertices shows that, with high probability, there exists a constant $\eta > 0$ such that every non-target vertex has an outgoing neighbour satisfying

$$\frac{\|X_j - q\|^2}{d} \leq \frac{\|X_i - q\|^2}{d} - \eta.$$

Since the initial normalized distance to the query is at most a constant, only $O(1)$ such decreases are required to reach the target region.

The following result separates two regimes for ρ_{crit}. For any fixed w, there is a constant $\tilde{c}(w)$. For $1 < c < \tilde{c}(w)$, a positive polynomial degree exponent is required. Once $c \geq \tilde{c}(w)$, the critical exponent vanishes. Note that $\rho_{\mathrm{crit}} = 0$ does not mean that the graph degree vanishes: We still need $n^{o(1)}$ degree to guarantee the improvement.

Lemma 1 (Informal version of Lemma 17)**.** *Fix $w \geq w_0$. There exists a $\tilde{c}(w)$ such that $\rho_{\mathrm{crit}}(c, w) = 0$ for all $c \geq \tilde{c}(w)$ and $\rho_{\mathrm{crit}}(c, w)$ is strictly decreasing for $1 < c < \tilde{c}(w)$. Moreover,*

$$\lim_{c \downarrow 1} \rho_{\mathrm{crit}}(c, w) = \frac{8}{9} + O(w^{-1/2}), \quad \tilde{c}(w) = 1 + \frac{1}{\sqrt{3w}} + O(w^{-1}).$$

In particular, for every fixed $c > 1$, $\rho_{\text{crit}}(c, w) = 0$ for all sufficiently large w.

Increasing the value of w shrinks the range of c that requires positive ρ_{crit}. However, as $c \to 1$, the required exponent remains non-trivial and approaches 8/9.

Finally, we show that the guarantees for greedy search on the τ-threshold graph also hold for the exact k-NN graph.

Lemma 2 (Informal version of Lemma 20)**.** *Let G_k be the directed KNN graph with $k = n^{\rho_k + o(1)}$, where $\rho_k > \rho_{\text{crit}}$. Choose ρ_t such that $\rho_{\text{crit}}(c, w) < \rho_t < \rho_k$. Then with high probability, $E(G_\tau(\rho_t)) \subseteq E(G_k)$. Thus, greedy search on G_k finds a c-approximate nearest neighbour of the query in $O(1)$ steps.*

4.2 Geometric Hubs and Greedy Navigation

In this subsection we work with the $G_\tau(\rho_s)$ for fixed $w \geq w_0$ and $0 < \rho_s < 1$. Our aim is to compare geometric hubs (vertices with high in-degree) and navigable hubs (vertices with high in-degree which are visited by the greedy search).

Definition 2 (Geometric Hub)**.** *A vertex in $G_\tau(\rho_s)$ is a geometric hub if its in-degree is higher than its out-degree.*

Definition 3 (Navigable Hub)**.** *A vertex v in $G_\tau(\rho_s)$ is a navigable hub if it is a geometric hub and there exists a vertex v' such that v is on the greedy path from v' to some query q with $\|q\|^2 = d$.*

Recall that in Section 3.2 we established that for all $1 \leq i \leq n$, $\|X_i\|^2/d$ lies inside the interval $[\alpha_{\min}, \alpha_{\max}]$ where $\alpha_{\min}$ and $\alpha_{\max}$ are constants depending only on w. In this section, we further establish the existence of constants α_{nav} and α_h such that

$$\alpha_{\min} < \alpha_{\text{nav}} < \alpha_h < 1 < \alpha_{\max},$$

and any vector v with $\alpha_v = \|v\|^2/d$, is a geometric hub if and only if $\alpha_v \in [\alpha_{\min}, \alpha_h]$ and v is a navigable hub if and only if $\alpha_v \in [\alpha_{\text{nav}}, \alpha_h]$. We will also show that the in-degree of v is a decreasing function of α_v. Combining these results establishes that the vertices with highest in-degree (those with $\alpha_v \in [\alpha_{\min}, \alpha_{\text{nav}}]$) are not visited during greedy search.

To establish the value of α_h, we again fix a vertex v, consider a random candidate $U \sim \mathcal{N}(0, I_d)$ and apply the LDP. In Section 3.2, we established an LDP for the event that U was an out-neighbour of v, while here we consider the event that U is an in-neighbour of v. The details are given in Appendix D.

Lemma 3. *[Informal version of Lemma 25] There is a function $\Delta(\alpha)$ with $\Delta(\alpha_h) = 0$ and $\Delta(\alpha)$ is strictly decreasing on α such that for v with $\|v\|^2/d = \alpha$,*

$$\frac{\mathbb{E}[D_{\mathbf{in}}(v)]}{\mathbb{E}[D_{\mathbf{out}}(v)]} = n^{\Delta(\alpha) + o(1)}.$$

The insight from this lemma is powerful. Previous results on hubness can be roughly summarized by the observation that vertices closer to the center tend to receive more incoming edges (Radovanovic et al., 2010; Angiulli, 2018). Lemma 3 makes this quantitative. At the boundary α_h, the incoming- and outgoing-degree exponents agree. Once we move a fixed amount towards the center, the exponent in Lemma 3 becomes positive, and the expected in-degree is polynomially larger in n, or equivalently exponentially larger in d. This explains why a relatively small change in norm can produce a very strong hubness effect. Above the boundary, the same exponent becomes negative, which gives the corresponding anti-hub behaviour discussed in Xu et al. (2026).

To study navigable hubs, we derive an LDP rate function to describe the event that v is selected at some step of the greedy search. We calculate α_{nav} to be the minimum value of α such that this rate function implies sufficiently high probability for such a vertex to exist. The details are given in Appendix D.

The relationship between $\alpha_{\min}$, α_{nav}, α_h, and $\alpha_{\max}$ offers a new perspective on the hubness phenomenon when compared with previous work. Rather than viewing hubs as a purely geometric phenomenon (Radovanovic et al., 2010; Xu et al., 2026; Angiulli, 2018), or a purely empirical observation (Munyampirwa et al., 2026), we have established a relationship between the geometry and the behaviour of greedy search. Perhaps surprisingly, we have shown that having a large in-degree does not imply that a vertex is useful for greedy navigation, even when the data is drawn from the Gaussian distribution.

4.3 Spherical Uniform Data and Comparison

In this section we further explore the importance of hubs by applying the same LDP analysis to data drawn from the spherical uniform distribution. Spherical uniform data 'looks like' Gaussian data in high dimensions (due to the concentration of the norm in Gaussian data) with the important distinction that it does not exhibit the hubness phenomenon.

We apply the same LDP analysis to the spherical uniform data and establish that the critical exponent for the k-NN graph is given by

$$\rho_{\text{uni}}(c,w) = 1 + \frac{w}{2}\log\left(1 - \left[z^2 + \sqrt{(1-z^2)\left(\left(1-\frac{s_{\text{uni,NN}}}{2}\right)^2 - z^2\right)}\right]^2\right)$$

where $z := 1 - \frac{c^2 s_{\text{uni,NN}}}{2}$ and

$$s_{\text{uni,NN}} = 2\left(1 - \sqrt{1-\exp\left(-2/w\right)}\right)$$

is the distance from q to its nearest neighbour. The details are given in Appendix E.1, Theorem 12.

This result is closely related to the work of Laarhoven (2018), who studies the intersections of spherical caps to characterize the search on uniform data. The main difference is that their analysis allows for restarting the greedy search if it gets stuck, while our result finds the critical degree required to make strict progress towards the query from every vertex in the graph.

Finally, we establish in the following theorem that there is a critical threshold $w_\flat$ in the moderate-high dimensional regime. For values of w below $w_\flat$, data from the spherical uniform distribution appears to be 'easier' to navigate, while for $w > w_\flat$, the Gaussian data is the easier case. This offers new evidence of the importance of hubs for nearest neighbour search on high dimensional data.

Theorem 2. *[Informal version of Theorem 13] For $w \geq w_0$,*

$$\lim_{c\downarrow 1}\rho_{\text{uni}}(c,w) = \frac{w}{2}\log\left(2 - e^{-2/w}\right).$$

At $w = w_0$, we have

$$\lim_{c\downarrow 1}\rho_{\text{crit}}(c,w) > \lim_{c\downarrow 1}\rho_{\text{uni}}(c,w).$$

On the other hand, there exists a constant $w_\flat$ such that for all $w > w_\flat$,

$$\lim_{c\downarrow 1}\rho_{\text{crit}}(c,w) < \lim_{c\downarrow 1}\rho_{\text{uni}}(c,w).$$

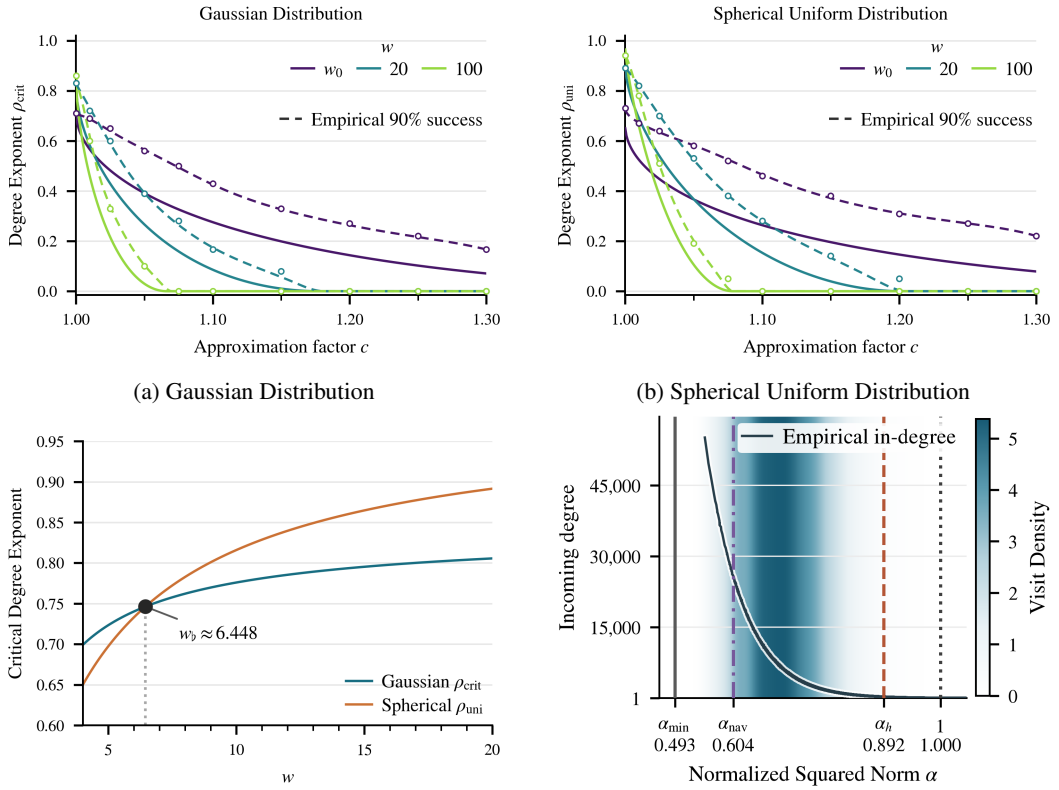
5 Numerical Results and Discussion

In this section, we provide some numerical results to illustrate and verify the theory developed in the paper. All numerical experiments use $n = 10^6$ points and exact directed k-NN graphs, with greedy search starting from uniformly sampled vertices.

Degree threshold $\rho_{\text{crit}}(c,w)$. The degree exponent thresholds $\rho_{\text{crit}}(c,w)$ and $\rho_{\text{uni}}(c,w)$ established above for the Gaussian and uniform distributions are difficult to study symbolically, and so Figures 1a and 1b plot them for a range of w and c values. The figures compare these theoretical values with the empirical degree exponent required to achieve a 90% recall for approximate nearest neighbour. While the empirical degrees are larger than the theory suggests, note that the theoretical values hold in the limit as $n \to \infty$, and the results show that the shape of the distributions match well.

Dimension threshold $w_\flat$. In Section 4.3 we established that there is a critical threshold $w_\flat$ such that for $d > w_\flat \log(n)$, nearest neighbour search on Gaussian data becomes easier than on spherical uniform data, suggesting that hubs accelerate the search. Figure 1c shows the theoretical critical degree exponents in the case where c is very close to 1, and establishes that $w_\flat \approx 6.448$.

Navigable hub threshold α_{nav}. In Section 4.2, we found that the range of hubs visited during the greedy search is a subset of the geometric hubs which have high degree. We derived an expression for α_{nav}, which is the minimum normalised squared norm of any point visited during the search. Figure 1d shows a density plot of the norms of visited vertices and we can see that the theoretical α_{nav} is a good approximation of the lower bound of this range. Moreover, the figure shows that there exist vertices with very high in-degree (those with norm between $\alpha_{\min}$ and α_{nav}) which are not visited by the greedy search, as predicted by the theory.



(a) Gaussian Distribution

(b) Spherical Uniform Distribution

(c) Near Exact Nearest Neighbour Search

(d) Comparison for Geometric hub and Navigable hub

Figure 1: Critical Degree Exponent and Navigation Comparisons

The insights in this paper open up several avenues for future work. The framework based on LDP developed here may be extended to more complex data distributions. For example, Gaussian mixture models would be a natural next step, since they provide a simple setting to study how greedy search navigates between different clusters. The distinction between geometric and navigable hubs may also suggest new search-oriented graph construction or searching strategies which take navigable hub into account.

AI USE STATEMENT

Generative AI tools were used for grammar checking and language polishing, for identifying potentially relevant related work, for sanity-checking mathematical proofs, and for assisting with visualization code. All mathematical arguments, code, and AI-assisted outputs were independently reviewed and verified by the authors. The authors take full responsibility for the final content of the paper.

REFERENCES

Fabrizio Angiulli. On the behavior of intrinsically high-dimensional spaces: distances, direct and reverse nearest neighbors, and hubness. *Journal of Machine Learning Research*, 18(170):1–60, 2018.

Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is "nearest neighbor" meaningful? In *International conference on database theory*, pp. 217–235. Springer, 1999.

Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38. Springer Science & Business Media, 2009.

Damien François, Vincent Wertz, and Michel Verleysen. The concentration of fractional distances. *IEEE Transactions on Knowledge and Data Engineering*, 19(7):873–886, 2007.

Cong Fu, Chao Xiang, Changxu Wang, and Deng Cai. Fast approximate nearest neighbor search with the navigating spreading-out graph. *Proceedings of the VLDB Endowment*, 12(5):461–474, 2019.

Chris R Giannella. Instability results for euclidean distance, nearest neighbor search on high dimensional gaussian data. *Information Processing Letters*, 169:106115, 2021.

Suhas Jayaram Subramanya, Fnu Devvrit, Harsha Vardhan Simhadri, Ravishankar Krishnawamy, and Rohan Kadekodi. Diskann: Fast accurate billion-point nearest neighbor search on a single node. *Advances in neural information processing Systems*, 32, 2019.

Thijs Laarhoven. Graph-Based Time-Space Trade-Offs for Approximate Near Neighbors. In *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99, pp. 57:1–57:14, 2018.

Wen Li, Ying Zhang, Yifang Sun, Wei Wang, Mingjie Li, Wenjie Zhang, and Xuemin Lin. Approximate nearest neighbor search on high dimensional data—experiments, analyses, and improvement. *IEEE Transactions on Knowledge and Data Engineering*, 32(8):1475–1488, 2019.

Yu A Malkov and Dmitry A Yashunin. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):824–836, 2018.

Blaise Munyampirwa, Vihan Lakshman, and Benjamin Coleman. Down with the hierarchy: The 'h' in hnsw stands for "hubs". In *European Conference on Information Retrieval*, pp. 33–48. Springer, 2026.

Liudmila Prokhorenkova and Aleksandr Shekhovtsov. Graph-based nearest neighbor search: From practice to theory. In *International Conference on Machine Learning*, pp. 7803–7813. PMLR, 2020.

Milos Radovanovic, Alexandros Nanopoulos, and Mirjana Ivanovic. Hubs in space: Popular nearest neighbors in high-dimensional data. *Journal of machine learning research*, 11(sept):2487–2531, 2010.

Anshumali Shrivastava, Zhao Song, and Zhaozhuo Xu. A theoretical analysis of nearest neighbor search on approximate near neighbor graph. *arXiv preprint arXiv:2303.06210*, 2023.

Mengzhao Wang, Xiaoliang Xu, Qiang Yue, and Yuxiang Wang. A comprehensive survey and experimental comparison of graph-based approximate nearest neighbor search. *Proceedings of the VLDB Endowment*, 14(11):1964–1978, 2021.

Xiaoliang Xu, Haonan Dai, Can Li, Mengzhao Wang, and Qiang Yue. Through the lens of hubness: A revisit on graph-based approximate nearest neighbor search:[experiments & analysis]. *Proceedings of the ACM on Management of Data*, 4(3 (SIGMOD):1–27, 2026.

A THRESHOLD MODEL AND PROBABILITY GEOMETRY

In this section we introduce a surrogate model for the KNN graph which we will use to make the analysis easier. Compared with previous theoretical work Prokhorenkova & Shekhovtsov (2020); Laarhoven (2018), since we work with the Gaussian rather than uniform distribution, we will need to spend a bit more time to justify that the surrogate model correctly reflects the KNN graph.

Section A.1 describes notation and preliminary theorems. Section A.2 builds the threshold model and shows that as the number of data points increases, it converges to the KNN graph in the limit. Section A.3 establishes the range of values that the norm of any data point might take, with high probability. This section also derives an expression for the distance between the query point and its nearest neighbor.

A.1 SETTING

We consider a data set

$$\mathcal{X} := (X_1, \ldots, X_n)$$

where

$$X_i \sim \mathcal{N}(0, I_d) \text{ for } i \in (1, \ldots, n), \quad \lim_{n\to\infty} \frac{d_n}{\log n} = w,$$

and w is a positive constant.

We will construct a graph on this dataset to model the directed KNN graph.

We follow Laarhoven (2018) in our parameterization of the graph degree. For some $\rho_s \in [0, 1)$, we define the target graph degree to be

$$k := n^{\rho_s + o(1)}.$$

Then we define

$$p := \frac{k}{n-1} = n^{\rho_s - 1 + o(1)} = \exp\left(-\frac{d(1-\rho_s)}{w} + o(d)\right) \tag{12}$$

to represent the desired probability that a point sampled from $\mathcal{N}(0, I_d)$ is a neighbor of a given vertex in our surrogate graph model.

Our surrogate graph model is constructed by connecting any given point v to every other point which lies within some specified threshold distance of v. In contrast with previous work, the threshold distance varies for different v. Specifically, it is a function of $\|v\|$.

Definition 4 (The τ-threshold graph)**.** *Let $\tau(\alpha)$ be the unique value such that*

$$\mathbb{P}\left[\frac{\|U - v\|^2}{d} \leq \tau(\alpha)\right] = \frac{k}{n-1},$$

when v is a fixed vector with $\|v\|^2 = d\alpha$, and $U \sim \mathcal{N}(0, I_d)$[3]. We define the τ-threshold graph on data set $\mathcal{X}$ to be a directed graph $G = (V, E)$ with vertex set $V := (X_1, ..., X_n)$. E contains the directed edge $X_i \to X_j$ if and only if

$$\|X_i - X_j\|^2 \leq \tau\left(\frac{\|X_i\|^2}{d}\right).$$

The analysis will focus on one step of the greedy search on the τ-threshold graph. Let $v \in \mathcal{X}$ be the data point corresponding to the current vertex (also named source) of the graph and let $q \in \mathbb{R}^d$ be the query point. We will use $U \sim \mathcal{N}(0, I_d)$ to represent another data point in $\mathcal{X}$. Throughout the analysis we will use

$$\alpha = \frac{\|v\|^2}{d}, \quad \beta = \frac{\|q\|^2}{d}$$

[3]Notice that by the rotational symmetry of $\mathcal{N}(0, I_d)$, the direction of v does not affect this probability

to represent the normalized lengths of v and q. We will assume, without loss of generality, that

$$v = \sqrt{\alpha d} e_1$$

and

$$q = \sqrt{\frac{d}{\alpha}}\theta e_1 + \sqrt{d\left(\beta - \frac{\theta^2}{\alpha}\right)} e_2,$$

where $e_1, \dots e_d$ are the standard basis vectors, and $\theta = \langle v, q \rangle d^{-1}$. Then, it will be useful to consider the projection of U onto the first two basis vectors, and the remaining dimensions:

$$A_d = \frac{\langle U, e_1 \rangle}{\sqrt{d}}, \quad B_d = \frac{\langle U, e_2 \rangle}{\sqrt{d}}, \quad M_d = \frac{1}{d}\sum_{j=3}^{d} \langle U, e_j \rangle^2.$$

In the analysis, we will consider the following three key distances, which we will sometimes refer to as the local term, greedy term, and current term respectively.

$$\begin{aligned}
\frac{\|U - v\|^2}{d} &= \frac{1}{d}\left[(U_1 - \sqrt{\alpha d})^2 + U_2^2 + \sum_{j=3}^{d} U_j^2\right] \\
&= (A_d - \sqrt{\alpha})^2 + B_d^2 + M_d\,. \\
\frac{\|U - q\|^2}{d} &= \frac{1}{d}\left[\left(U_1 - \theta\sqrt{\frac{d}{\alpha}}\right)^2 + \left(U_2 - \sqrt{d\left(\beta - \frac{\theta^2}{\alpha}\right)}\right)^2 + \sum_{j=3}^{d} U_j^2\right] \\
&= \left(A_d - \frac{\theta}{\sqrt{\alpha}}\right)^2 + \left(B_d - \sqrt{\beta - \frac{\theta^2}{\alpha}}\right)^2 + M_d\,. \\
\frac{\|q - v\|^2}{d} &= \alpha + \beta - 2\theta\,.
\end{aligned}$$

Throughout the analysis, we will make use of the large deviation principle, and related results. Dembo and Zeitouni Dembo & Zeitouni (2009) provide a useful reference for applying this technique, and we reproduce the key results for our application below.

Theorem 3 (Gärtner-Ellis Theorem (Dembo & Zeitouni (2009), Theorem 2.3.6))**.** *Consider a sequence of random vectors $Y_d \in \mathbb{R}^l$. where Y_d possesses the law μ_d and log-MGF*

$$\Lambda_d(\lambda) \triangleq \log \mathbb{E}\left[e^{\langle \lambda, Y_d \rangle}\right].$$

If, for all $\lambda \in \mathbb{R}^l$, the limit of the log-MGF

$$\Lambda(\lambda) \triangleq \lim_{d\to\infty} \frac{1}{d}\Lambda_d(d\lambda)$$

exists as an extended real number, 0_l belongs to the interior of

$$D_\Lambda \triangleq \{\lambda \in \mathbb{R}^l : \Lambda(\lambda) < \infty\},$$

and Λ is an essentially smooth, lower semicontinuos function, then the LDP holds with speed d and good rate function $\Lambda^*(\cdot)$ where $\Lambda^*(\cdot)$ is the Fenchel-Legendre transformation of $\Lambda(\cdot)$ with

$$D_{\Lambda^*} = \{x \in \mathbb{R}^l : \Lambda^*(x) < \infty\}.$$

Remark 3. *If $\{\mu_d\}$ satisfies an LDP with speed d and rate function Λ^*. then*

$$-\inf_{x\in F^\circ} \Lambda^*(x) \le \liminf_{d\to\infty} \frac{1}{d}\log \mu_d(F) \le \limsup_{d\to\infty} \frac{1}{d}\log \mu_d(F) \le -\inf_{x\in \bar{F}} \Lambda^*(x).$$

We can obtain the limit

$$\lim_{d\to\infty} \frac{1}{d} \log \mu_n(F) = - \inf_{x\in F} \Lambda^*(x)$$

if F is a Λ^*-continuity set, meaning that

$$\inf_{x\in F^\circ} \Lambda^*(x) = \inf_{x\in \bar{F}} \Lambda^*(x).$$

Theorem 4 (Contraction Principle (Dembo & Zeitouni (2009), Theorem 4.2.1))**.** *Let X and Y be Hausdorff topological spaces and*

$$f : X \to Y$$

be a continuous function. Consider a good rate function $I : X \to [0, \infty]$.

(a) for each $y \in Y$, define:

$$I'(y) \triangleq \inf\{I(x) : x \in X,\ y = f(x)\}.$$

then I' is a good rate function on Y, where as usual the infimum over the empty set is taken as ∞.

(b) If I controls the LDP associated with a family of probability measures $\{\mu_\varepsilon\}$ on X, then I' controls the LDP associated with family of probability measures $\{\mu_\varepsilon \circ f^{-1}\}$ on Y.

We will use this theorem with X and Y equal to the set of real numbers, which is a Hausdorff topological space. Through a standard application of Theorem 3, we will show that the LDP holds for A_d, B_d, and M_d which follow the standard normal distribution and chi-squared distributions. Then, we will use the contraction principle to analyze the distribution of neighbors of the vertex v in order to show that we can always find a neighbor which is closer to the query than the current vertex.

The current state of the greedy search is entirely captured by the 3-tuple $S = (\alpha, \beta, \theta)$. Now, any candidate neighbor is defined by the random variables A_d, B_d, and M_d. For any corresponding realisation of these variables, a, b, and m, we define

$$g_l(a, b, m; S) = (a - \sqrt{\alpha})^2 + b^2 + m\,,$$
$$g_G(a, b, m; S) = \left(a - \frac{\theta}{\sqrt{\alpha}}\right)^2 + \left(b - \sqrt{\beta - \frac{\theta^2}{\alpha}}\right)^2 + m\,,$$

to be the distance from the candidate to the current vertex and the query vertex respectively.

To complete the proof we must show that there is a candidate (a, b, m) which satisfies three requirements: it must have enough probability mass to occur, be close enough to the current vertex, and closer to the query than current vertex. If these three requirements hold, it will imply that we can make progress towards the query.

A.2 Threshold Model

In this subsection we develop some probabilistic tools based on the LDP for the threshold model which will be used in the remainder of the analysis. In the end we show that the τ-threshold graph in Definition 4 correctly reflects the behavior of the KNN graph, and derive an expression for the threshold τ.

We will start by showing that the distribution of points in the dataset satisfies the large deviation principle.

Theorem 5. *The law of (A_d, B_d, M_d) satisfies an LDP on $\mathbb{R}^3$ with speed d and good rate function*

$$J(a, b, m) := \begin{cases} \frac{1}{2}(a^2 + b^2 + m - 1 - \log m) & \text{if } m > 0 \\ \infty & \text{if } m \le 0. \end{cases}$$

Proof. In order to apply Theorem 3, the proof consists of the following three steps.

1. The limit of the log-MGF $\Lambda(\cdot)$ exists, and the origin belongs to its domain D_Λ.

2. The limit of the log-MGF is a smooth, lower semicontinuous function.

3. The Fenchel-Legendre transformation of $\Lambda(\cdot)$ is the function $J(\cdot)$ given in the theorem statement.

Step 1: Notice, that $\sqrt{d}A_d$ and $\sqrt{d}B_d$ are distributed according to the standard normal distribution $N(0,1)$ and that dM_d is distributed according to the chi-squared distribution χ^2_{d-2}. Then, the scaled log moment generating function when $\lambda_m < 1/2$ is given by

$$\begin{aligned}\Lambda_d(d\lambda_a, d\lambda_b, d\lambda_m) &:= \log \mathbb{E} \exp\{d(\lambda_a A_d + \lambda_b B_d + \lambda_m M_d)\} \\ &= \log \mathbb{E} e^{d\lambda_a A_d} + \log \mathbb{E} e^{d\lambda_b B_d} + \log \mathbb{E} e^{d\lambda_m M_d} \quad \text{(by independence)} \\ &= \frac{d\lambda_a^2}{2} + \frac{d\lambda_b^2}{2} + \log\left((1-2\lambda_m)^{-(d-2)/2}\right) \quad \text{(by MGF of } \mathcal{N} \text{ and } \chi^2_{d-2}) \\ &= \frac{d\lambda_a^2}{2} + \frac{d\lambda_b^2}{2} - \frac{d-2}{2}\log\left(1-2\lambda_m\right).\end{aligned}$$

When $\lambda_m \geq 1/2$, the MGF of χ^2_{d-2} diverges and $\Lambda_d(\lambda_a, \lambda_b, \lambda_m) = \infty$. Taking the limit as $d \to \infty$ gives

$$\Lambda(\lambda_a, \lambda_b, \lambda_m) = \lim_{d\to\infty} \frac{1}{d}\Lambda_d(d\lambda_a, d\lambda_b, d\lambda_m) = \begin{cases} \frac{1}{2}\left[\lambda_a^2 + \lambda_b^2 - \log(1-2\lambda_m)\right] & \text{if } \lambda_m < \frac{1}{2} \\ +\infty & \text{if } \lambda_m \geq \frac{1}{2} \end{cases}$$

Thus the limit exists and the effective domain is

$$\mathcal{D}_\Lambda := \mathbb{R}^2 \times \left(-\infty, \frac{1}{2}\right),$$

which contains the origin.

Step 2: Let us examine the properties of $\Lambda(\lambda_a, \lambda_b, \lambda_m)$.

1. The domain $\mathcal{D}_\Lambda^o$ is non-empty.

2. $\nabla\Lambda(\lambda_a, \lambda_b, \lambda_m) = \left(\lambda_a, \lambda_b, \frac{1}{1-2\lambda_m}\right)$ which means it is differentiable throughout $\mathcal{D}_\Lambda^o$.

3. The boundary of $\mathcal{D}_\Lambda^o$ is $\mathbb{R}^2 \times \{\frac{1}{2}\}$. Since $1 - 2\lambda_m \to 0$ when $\lambda_m \uparrow \frac{1}{2}$, it holds that $\lim_{d\to\infty} \|\nabla\lambda\| = \infty$ when $\lambda_m \uparrow \frac{1}{2}$.

Combining these properties establishes that Λ is essentially smooth.

Finally, since Λ is continuous inside $\mathcal{D}_\Lambda$, equals $+\infty$ outside $\mathcal{D}_\Lambda$, and diverges to $+\infty$ at the boundary $\lambda_m = 1/2$, it holds that Λ is lower-semicontinuous.

Step 3: Now we can apply the Legendre transformation to obtain the good rate function. Recall that the Legendre transformation is given by

$$\Lambda^*(a,b,m) = \sup_{\substack{\lambda_a,\lambda_b\in\mathbb{R} \\ \lambda_m < \frac{1}{2}}} \left\{a\lambda_a + b\lambda_b + m\lambda_m - \Lambda(\lambda_a, \lambda_b, \lambda_m)\right\}.$$

Substituting Λ, we have

$$\Lambda^*(a,b,m) = \sup_{\lambda_a\in\mathbb{R}}\left\{a\lambda_a - \frac{1}{2}\lambda_a^2\right\} + \sup_{\lambda_b\in\mathbb{R}}\left\{b\lambda_b - \frac{1}{2}\lambda_b^2\right\} + \sup_{\lambda_m<1/2}\left\{m\lambda_m + \frac{1}{2}\log(1-2\lambda_m)\right\}.$$

For the first two terms, we have that the derivative of $cx - x^2/2)$ with respect to x is $(a-x)$, and the second derivative is negative. Thus the maximum is achieved when $x = a$. This corresponds to $\lambda_a = a$ and $\lambda_b = b$.

For the final term, we have that the derivative of $cx + (1/2)\log(1-2x)$ with respect to x is $c - (1-2x)^{-1}$, and the second derivative is negative. Thus, the maximum is achieved when

$x = \frac{1}{2} - \frac{1}{2c}$. This corresponds to $\lambda_m = \frac{1}{2} - \frac{1}{2m}$. Since the domain of Λ requires that $\lambda_m < \frac{1}{2}$, we require that $m > 0$.

Combining these results, we have

$$\Lambda^*(a, b, m) = \frac{1}{2}\left(a^2 + b^2 + m - 1 - \log(m)\right).$$

Then, applying Theorem 3 completes the proof.

□

Now, it is not difficult to show that other quantities of interest also satisfy an LDP. In the next lemma, we show that the distribution of distances to the current vertex satisfies an LDP.

Theorem 6. *For every fixed $\alpha \geq 0$, the law of $Z_d = \frac{\|U-v\|^2}{d}$ satisfies an LDP with speed d and good rate function*

$$I_\alpha(x) = \begin{cases} \frac{1}{2}(m_\alpha(x) - 1 - \log m_\alpha(x) + \alpha(m_\alpha(x) - 1)^2) & \text{if } x > 0 \\ +\infty & \text{if } x \leq 0, \end{cases}$$

where, for $x > 0$,

$$m_\alpha(x) = \frac{2x}{1 + \sqrt{1 + 4\alpha x}} \text{ is the positive solution of } \alpha m^2 + m = x.$$

Proof. Recall that we can write the random variable Z_d using the function $g_l(a, b, m) = (a - \sqrt{\alpha})^2 + b^2 + m$. The map

$$g_l : \mathbb{R}^3 \to \mathbb{R}$$

is continuous. By Theorem 5 we know (A_d, B_d, M_d) satisfies an LDP with speed d and good rate function $J(a, b, m)$. By Theorem 4, we have that the rate function of Z_d is

$$I_\alpha(x) = \inf\{J(a, b, m) : g_l(a, b, m) = x\}$$

We will solve this optimization problem by considering two cases.

Case 1: $x \leq 0$. By definition, it is always the case that $g_l(a, b, m) > 0$ and so there is no solution for $g_l(a, b, m) = x \leq 0$. Thus in this case, $I_\alpha(x) = \inf\{\emptyset\} = +\infty$.

Case 2: $x > 0$. The feasible set is $\{(a, b, m) : m > 0,\ g_l(a, b, m) = x\}$, which is non-empty. It is clear to see that $\nabla^2 J(a, b, m) \succ 0$, and so $J(a, b, m)$ is strictly convex and the stationary point is its unique global minimizer. To solve the contrained optimization problem, we can use the Lagrange Multiplier

$$\mathcal{L}(a, b, m, \lambda_\mathcal{L}) = \frac{1}{2}(a^2 + b^2 + m - 1 - \log m) - \lambda_\mathcal{L}\left((a - \sqrt{\alpha})^2 + b^2 + m - x\right).$$

Taking the derivative for each term and setting them equal to 0, we obtain the following system of equations.

$$\begin{gathered} a - 2\lambda_\mathcal{L}(a - \sqrt{\alpha}) = 0 \\ b(1 - 2\lambda_\mathcal{L}) = 0 \\ \lambda_\mathcal{L} = \frac{1}{2} - \frac{1}{2m} \\ (a - \sqrt{\alpha})^2 + b^2 + m = x \end{gathered}$$

Notice that there is no solution when $b \neq 0$ due to the fact that $1/m$ cannot be 0. Thus we have that $b = 0$. From the third equation, we can establish that $2\lambda_\mathcal{L} = 1 - m^{-1}$. Substituting and solving the first equation establishes that $a = (1 - m)\sqrt{\alpha}$. Substituting everything into the final equation gives

$$\alpha m^2 + m - x = 0.$$

Thus, the minimum is achieved then $a = (1 - m_\alpha(x))\sqrt{(}\alpha)$, $b = 0$, and $m = m_\alpha(x)$, where

$$m_\alpha(x) = \frac{-1 + \sqrt{1 + 4\alpha x}}{2\alpha} = \frac{2x}{1 + \sqrt{1 + 4\alpha x}}$$

is the positive solution of $\alpha m^2 + m = x$. Substituting this minimizer back into the definition of J gives the claimed expression and completes the proof. □

We now prove some properties of the rate function of Z_d.

Lemma 4. *The function I_α satisfies*

$$I_\alpha(1+\alpha) = 0, \quad I_\alpha(s) \to +\infty \quad \text{as } s \downarrow 0,$$

and I_α is continuous and strictly decreasing on $(0, 1+\alpha)$.

Proof. For $s > 0$, write

$$m = m_\alpha(s), \qquad s = \alpha m^2 + m.$$

Then $s = 1 + \alpha$ if and only if $m = 1$. Therefore

$$I_\alpha(1+\alpha) = \frac{1}{2}(1 - 1 - \log 1 + \alpha(1-1)^2) = 0.$$

Also, as $s \downarrow 0$, we have $m \downarrow 0$, and hence $-\log m \to +\infty$. Thus

$$I_\alpha(s) \to +\infty \qquad \text{as } s \downarrow 0.$$

The function $m_\alpha(s)$ is continuous and strictly increasing in s, because $s = \alpha m^2 + m$ is strictly increasing in $m \geq 0$.

For $m \in (0, 1)$,

$$\frac{d}{dm} I_\alpha(\alpha m^2 + m) = \frac{1}{2}\left(1 - \frac{1}{m} + 2\alpha(m-1)\right) = \frac{1}{2}\left(\frac{1}{m} + 2\alpha\right)(m-1) < 0.$$

Therefore I_α is strictly decreasing on $(0, 1+\alpha)$. □

We can now define the asymptotic threshold distance, $s_*(\alpha)$ of the surrogate graph model for a data point which has norm $\sqrt{d\alpha}$ in terms of this rate function.

Definition 5 (Threshold distance s_*)**.** *Fix $0 < \rho_s < 1$. For every $\alpha \geq 0$, let $s_*(\alpha; \rho_s) \in (0, 1+\alpha)$ be the unique solution of*

$$I_\alpha(s_*(\alpha; \rho_s)) = \frac{1-\rho_s}{w}.$$

We will write $s_*(\alpha) = s_*(\alpha; \rho_s)$ when ρ_s is clear from the context. Note that the existence and uniqueness of the solution follow from Lemma 4.

In the next theorem, we see that using $s_*(\alpha)$ as the threshold function for a τ-threshold graph corresponds asymptotically to the KNN graph with $k = n^{\rho_s + o(1)}$.

Theorem 7. *For a fixed current state v with $\alpha = \frac{\|v\|^2}{d}$,*

$$\mathbb{P}\left[\frac{\|U - v\|^2}{d} \leq s_*(\alpha)\right] = \exp\left(-d\frac{1-\rho_s}{w} + o(d)\right) = \frac{n^{\rho_s + o(1)}}{n-1}.$$

Proof. By the monotonicity of I_α we have

$$\inf_{s \leq s_*(\alpha)} I_\alpha(s) = I_\alpha(s_*(\alpha)).$$

Moreover, since I_α is continuous, $(-\infty, s_*(\alpha)]$ is an I_α-continuity set. Then the LDP gives (see Remark 3)

$$\lim_{d\to\infty} \frac{1}{d} \log \mathbb{P}(Z_d \leq s_*(\alpha)) = -I_\alpha(s_*(\alpha)) = -\frac{1-\rho_s}{w}.$$

From this, we have

$$\begin{aligned}
&\frac{1}{d}\log \mathbb{P}(Z_d \leq s_*(\alpha)) = -\frac{1-\rho_s}{w} + o(1) \\
\Rightarrow \quad &\log \mathbb{P}(Z_d \leq s_*(\alpha)) = -\frac{d(1-\rho_s)}{w} + o(d) \\
\Rightarrow \quad &\mathbb{P}(Z_d \leq s_*(\alpha)) = \exp\left(\frac{-d(1-\rho_s)}{w} + o(d)\right),
\end{aligned}$$

which completes the proof by equation (12). □

In fact, we can show an even sharper bound which guarantees with high probability that high degrees do not happen.

Lemma 5. *For every fixed $0 < \rho_s < 1$, with probability $1 - o(1)$,*

$$\max_{i \le n} D_{out}(X_i) \le n^{\rho_s + o(1)}.$$

Proof. Let

$$p = \mathbb{P}\left[\frac{\|U - v\|^2}{d} \le s_*(\alpha)\right] = \exp\left(-\frac{d(1-\rho_s)}{w}\right)$$

The degree of any fixed vertex v then follows a binomial distribution:

$$D_{\text{out}}(v) \sim \text{Binomial}(n-1, p)$$

The expectation is

$$\bar{\mu} \triangleq \mathbb{E}[D_{\text{out}}(v)] = (n-1)p = n^{\rho_s + o(1)}.$$

By the Chernoff bound, we have that

$$\mathbb{P}\left(D_{\text{out}}(X_i) \ge 2\bar{\mu}\right) \le \exp\left(-\frac{\bar{\mu}}{3}\right).$$

Then by the union bound,

$$\mathbb{P}\left(\max_{i \le n} D_{\text{out}}(X_i) \ge 2\bar{\mu}\right) \le n \exp\left(-\frac{1}{3}\bar{\mu}\right) \to 0,$$

which completes the proof. □

A.3 Norm Interval and Nearest Neighbour Scale

Although we treat v as a fixed current point in our one step analysis, we still need to consider the range of possible values of $\|v\|$. In this section, we derive an expression for the maximum and minimum values of α. We also define in this section the nearest neighbor distance scale. That is, the distance we expect to see between the query point and its nearest neighbor.

We begin this section be establishing an LDP on the norm of X_i.

Lemma 6. *The law of $z_i = \frac{\|X_i\|^2}{d}$ satisfies an LDP on $\mathbb{R}$ with speed d with the good rate function*

$$I_0(x) := \begin{cases} \frac{1}{2}(x - 1 - \log(x)) & \text{if } x > 0 \\ +\infty & \text{if } x \le 0 \end{cases}$$

Proof. Notice that the distribution of z_i is equal to the distribution of Z_d in Theorem 6 when $\alpha = 0$. Thus, this lemma is a special case of Theorem 6. When $\alpha = 0$, we have that $m_\alpha(x) = x$, from which the result follows immediately. □

We can now derive expressions for the range of the norms of the data points.

Lemma 7. *Let $\alpha_{\min}$ and $\alpha_{\max}$ be the solutions of*

$$\frac{1}{2}(\alpha - 1 - \log(\alpha)) = \frac{1}{w}.$$

Then

$$\mathbb{P}\left[\frac{\|X_i\|^2}{d} \in [\alpha_{\min}, \alpha_{\max}]\right] \to 1\,.$$

Moreover, for any constant $h > 0$, we have that

$$\mathbb{P}\left[\frac{\|X_i\|^2}{d} \in [\alpha_{\min} - h, \alpha_{\max} + h] \;\; \textit{for every } i \le n\right] = 1 - o(1)$$

Proof. Let $K := (0, \alpha_{min}) \cup (\alpha_{max}, \infty)$. By the LDP (see Remark 3), we have

$$\limsup_{d\to\infty} \frac{1}{d} \log\left(\mathbb{P}\left[\frac{\|X_i\|}{d} \in K\right]\right) \leq -\inf_{\overline{K}} \frac{1}{2}(\alpha - 1 - \log(\alpha))$$

The derivative

$$\frac{d}{d\alpha}\, \alpha - 1 - \log(\alpha) = 1 - \frac{1}{\alpha}$$

is 0 when $\alpha = 1$. Now we have

$$\inf_{\overline{K}} \frac{1}{2}(\alpha - 1 - \log(\alpha)) = \min\left\{\frac{1}{2}(\alpha_{min} - 1 - \log(\alpha_{min})), \frac{1}{2}(\alpha_{max} - 1 - \log(\alpha_{max}))\right\}.$$

By definition we have $\frac{1}{2}(\alpha_{min} - 1 - \log(\alpha_{min})) = \frac{1}{2}(\alpha_{max} - 1 - \log(\alpha_{max})) = \frac{1}{w}$. Thus

$$\mathbb{P}\left[\frac{\|X_i\|^2}{d} \in K\right] \leq \exp\left(-\frac{d}{w} + o(d)\right) = \exp(-\log n + o(d)) = n^{-1+o(1)} \to 0\,.$$

Which proves the first claim of the lemma. For the universal event, notice that

$$\mathbb{E}\left[\frac{\|X_i\|^2}{d}\right] = 1, \quad \mathrm{Var}\left[\frac{\|X_i\|^2}{d}\right] = \frac{2}{d},$$

from which Chebyshev's inequality gives the claim immediately. □

Notice that the definition of $\alpha\min$ and $\alpha_{\max}$ depend only on w, which is a constant, and thus $\alpha_{\min}$ and $\alpha_{\max}$ can also be thought of as constants.

Definition 6. *let $m_{\gamma,\mathrm{NN}} \in (0,1)$ be the solution to*

$$\frac{1}{2}(m_{\gamma,\mathrm{NN}} - 1 - \log(m_{\gamma,\mathrm{NN}}) + \gamma(m_{\gamma,\mathrm{NN}} - 1)^2) = \frac{1}{w}\,.$$

and let

$$s_{\mathrm{NN}}(\gamma) = \gamma m_{\gamma,\mathrm{NN}}^2 + m_{\gamma,\mathrm{NN}}\,.$$

The following lemma shows that for any target query q with $\beta = \|q\|^2/d$, the nearest neighbor of q will be at distance $s_{\mathrm{NN}}(\beta)$. Then for some $c \geq 1$, we can say that the c-ANN search is complete if and only if

$$\alpha + \beta - 2\theta \leq c \cdot s_{NN}(\beta).$$

Lemma 8. *It holds that*

$$\frac{\min_{i\leq n} \|X_i - q\|^2/d}{s_{\mathrm{NN}}(\beta)} \to 1.$$

Proof. By considering q to be the current state, we can use Theorem 6 directly. From Definition 6 we have

$$I_\beta(s_{\mathrm{NN}}(\beta)) = \frac{1}{w}.$$

Now fix any $\epsilon > 0$ and let

$$s_- = (1-\epsilon)s_{NN}(\beta), \quad s_+ = (1+\epsilon)s_{NN}(\beta).$$

By Lemma 4 we have

$$I_\beta(s_-) > \frac{1}{w}, \quad I_\beta(s_+) < \frac{1}{w}.$$

Now consider the lower bound of such nearest-neighbour distance, we have

$$\begin{aligned}
\mathbb{P}\left(\min_{i\leq n} \|X_i - q\|^2/d \leq s_-\right) &\leq n\mathbb{P}\left(\|X_i - q\|^2/d \leq s_-\right) \\
&= n\exp\left(-I_\beta(s_-)\right) \\
&= \exp\left(d\left(\frac{1}{w} - I_\beta(s_-)\right) + o(d)\right) \to 0,
\end{aligned}$$

where we use the fact that $\log(n) = \frac{d}{w} + o(d)$. For the upper bound, since $I_\beta(s_+) < \frac{1}{w}$ we have

$$n\mathbb{P}\left(\|X_i - q\|^2/d \le s_+\right) = \exp\left(d\left(\frac{1}{w} - I_\beta(s_+)\right) + o(d)\right) \to \infty.$$

By independence,

$$\mathbb{P}\left(\min_{i\le n} \|X_i - q\|^2/d \ge s_+\right) = \left(1 - \mathbb{P}\left(\|X_i - q\|^2/d \le s_+\right)\right)^n \le \exp\left(-n\mathbb{P}\left(\|X_i - q\|^2/d \le s_+\right)\right) \to 0.$$

Hence, with probability tending to one,

$$(1-\epsilon)s_{NN}(\beta) < \min_{i\le n} \|X_i - q\|^2/d < (1+\epsilon)s_{NN}(\beta).$$

As $\epsilon > 0$ is arbitrary, the theorem holds. □

The next lemma provides a general method for determining whether there is a data point in a given region.

Lemma 9. *Let*

$$Y_i := (A_{d,i}, B_{d,i}, M_{d,i}),$$

for $i = 1, \cdots, n$ be independent candidates, and let $B \subset \mathbb{R}^3$ be a J-continuity set. Write

$$I_B := \inf_{(a,b,m)\in B} J(a,b,m).$$

Then,

- *if $I_B < \frac{1}{w}$, then $\mathbb{P}(\exists i \le n : Y_i \in B) \to 1$, and*
- *if $I_B > \frac{1}{w}$, then $\mathbb{P}(\exists i \le n : Y_i \in B) \to 0$.*

Proof. By Theorem 5 we have

$$p := \mathbb{P}(Y_1 \in B) = \exp(-dI_B + o(d)).$$

Let $N_B := |\{i : Y_i \in B\}|$. Then, $N_B \sim \text{Binomial}(n, p)$. Hence, since $\log(n) = d/w + o(d)$,

$$\mathbb{E}(N_B) = n \cdot p = \exp\left(-d\left[I_B - \frac{1}{w}\right] + o(d)\right).$$

Thus, if $I_B < \frac{1}{w}$ then $\mathbb{E}(N_B) \to \infty$, and

$$\mathbb{P}(N_B = 0) = (1-p)^n \le \exp(-np) \to 0\,,$$

and if $I_B > \frac{1}{w}$ then $\mathbb{E}(N_B) \to 0$, and

$$\mathbb{P}(N_B \ge 1) \le \mathbb{E}(N_B) \to 0\,.$$

□

B Navigate Family and The Local Radius Required for progress

Throughout this section we use $\beta = 1$, since we consider only a single query point and this equality will hold with high probability in the limit. The norm of the vector at the current iteration of the greedy search may vary over the full interval $[\alpha_{min}, \alpha_{max}]$. The current state of the greedy search can be expressed by $S = (\alpha, \beta, \theta)$, where θ is the normalized inner product between the current vector and the query. We will additionally represent the normalized distance between the current vector and the query by $r = \alpha + \beta - 2\theta$.

Given the current vector v, the greedy search will minimize the distance to the query $g_G = \frac{\|U-q\|^2}{d}$ among all outgoing neighbors. To consider its behavior we first solve the question in continuous space. Let $x = (a, b, m)$ with $m > 0$ be a continuous candidate. Lemma 9 establishes that if $J(x) \le 1/w$,

then there will be a point arbitrarily close to x in our dataset with high probability. Moreover, if the distance from x to the current vector is less than $s_*(\alpha; \rho_s)$, then x will be a neighbor of the current vertex. So, we would like to find

$$\min_{x \in \mathbb{R}^3} \left\{ g_G(x; S) : J(x) \leq \frac{1}{w}, g_l(x; S) \leq s_*(\alpha; \rho_s) \right\}.$$

In order to make progress towards the query, we need to know the conditions under which this minimum will be less than the current term r. In short, we would like to pick a ρ_s such that we can make progress whenever $r > c \cdot s_{\text{NN}}$. The proof of Theorem 7 establishes that $\rho_s = 1 - wI\left(s_*(\alpha; \rho_s)\right)$. Thus, minimizing ρ_* is equivalent to minimizing the threshold distance $s_*(\alpha; \rho_s)$, which gives rise to the following natural reverse problem

$$\min_{x \in \mathbb{R}^3} g_l(x; S) \quad \text{subject to } J(x) \leq \frac{1}{w} \quad g_G(x; S) \leq r.$$

At useful transformation of the optimization problem is to parameterize it by $\lambda \in [0, 1]$ and rewrite it as

$$\min_x [\lambda g_l(x; S) + (1 - \lambda) g_G(x; S)] \quad \text{s.t. } J(x) \leq \frac{1}{w}. \qquad (*)$$

Then if x_λ solves $(*)$, let $r_\lambda = g_G(x_\lambda; S)$. Then x_λ also solves

$$\min g_l(x; S) \quad \text{subject to } J(x) \leq \frac{1}{w} \quad g_G(x; S) \leq r_\lambda.$$

B.1 Navigate Family and One Step Critical Threshold

Lemma 10. *Let $C_\lambda := \lambda v + (1 - \lambda) q$. Then, the following optimization problems are equivalent.*

$$\min \left[\lambda \frac{\|U - v\|^2}{d} + (1 - \lambda) \frac{\|U - q\|^2}{d} \right] \iff \min \left[\frac{\|U - C_\lambda\|^2}{d} \right].$$

Proof. For an arbitrary vector U,

$$\begin{aligned}
\lambda \frac{\|U - v\|^2}{d} + (1 - \lambda) \frac{\|U - q\|^2}{d} &= \frac{1}{d} \Big[\lambda(\|U\|^2 - 2\langle U, v \rangle + \|v\|^2) + \\
&\qquad (1 - \lambda)(\|U\|^2 - 2\langle U, q \rangle + \|q\|^2) \Big] \\
&= \frac{1}{d} \big[\|U\|^2 - 2(\langle U, \lambda v + (1 - \lambda) q \rangle) + \|\lambda v + (1 - \lambda) q\|^2 \\
&\qquad - \|\lambda v + (1 - \lambda) q\|^2 + \lambda \|v\|^2 + (1 - \lambda) \|q\|^2 \big] \\
&= \frac{\|U - (\lambda v + (1 - \lambda) q)\|^2}{d} + \lambda (1 - \lambda) \frac{\|v - q\|^2}{d}
\end{aligned}$$

The first term on the final line is exactly $\frac{\|U - C_\lambda\|^2}{d}$ and the second term is fixed, which completes the proof. □

Theorem 8. *For a fixed $\lambda \in [0, 1]$, define the normalized squared norm of C_λ as*

$$\alpha_\lambda := \frac{\|C_\lambda\|^2}{d} = \lambda \alpha + (1 - \lambda) - \lambda (1 - \lambda) r.$$

Let $m_{\alpha_\lambda, \text{NN}} \in (0, 1)$ be the unique solution of

$$\frac{1}{2}(m_{\alpha_\lambda, \text{NN}} - 1 - \log(m_{\alpha_\lambda, \text{NN}}) + \alpha_\lambda (m_{\alpha_\lambda, \text{NN}} - 1)^2) = \frac{1}{w}.$$

Let $m_\lambda = m_{\alpha_\lambda, \text{NN}}$, and define

$$x_\lambda := (a_\lambda, b_\lambda, m_\lambda),$$

where

$$a_\lambda := (1-m_\lambda)\left(\lambda\sqrt{\alpha}+(1-\lambda)\frac{\theta}{\sqrt{\alpha}}\right), \quad b_\lambda := (1-m_\lambda)(1-\lambda)\sqrt{1-\frac{\theta^2}{\alpha}}.$$

Then x_λ is the unique solution of

$$\min_x[\lambda g_l(x;S)+(1-\lambda)g_G(x;S)] \quad \textit{s.t. } J(x)\le\frac{1}{w}.$$

It gives the minimum value

$$s_\lambda = \alpha_\lambda m_\lambda^2 + m_\lambda + \lambda(1-\lambda)r.$$

Proof. By Lemma 10, we have that the minimization problem in the theorem statement is equivalent to

$$\min_x\left[\frac{\|U-C_\lambda\|^2}{d}\right] \quad \text{s.t. } J(x)\le\frac{1}{w}.$$

We can reuse the methods from Theorem 6, Lemma 4, and Theorem 7 to solve this. Firstly, the norm of C_λ is given by

$$\begin{aligned}\frac{\|C_\lambda\|^2}{d} &= \lambda^2\frac{\|v\|^2}{d}+(1-\lambda)^2\frac{\|q\|^2}{d}+2\lambda(1-\lambda)\frac{\langle v,q\rangle}{d}\\ &= \lambda^2\alpha+(1-\lambda)^2+2\lambda(1-\lambda)\theta.\end{aligned}$$

Recall that

$$r=\alpha+1-2\theta,$$

which gives $\alpha_\lambda = \lambda\alpha+(1-\lambda)-\lambda(1-\lambda)r$.

Then by Theorem 6 for a source that has norm α_λ the rate function of $\frac{\|U-C_\lambda\|^2}{d}$ is $I_{\alpha_\lambda}(x)$.

By Lemma 4, this rate function is monotonically decreasing on $(0,1+\alpha_\lambda)$ and the smallest distance x that achieves

$$J(x)\le\frac{1}{w}$$

is the unique solution of $I_{\alpha_\lambda}(x)=1/w$. Then the minimizer x_λ is obtained from same Lagrange multiplier method in Theorem 6.

Lemma 10 gives the minimum value

$$s_\lambda = s_{NN}(\alpha_\lambda)+\lambda(1-\lambda)r = \alpha_\lambda m_\lambda^2+m_\lambda+\lambda(1-\lambda)r.$$

□

Lemma 11. *At a fixed feasible state with $r>0$, the navigate family gives*

$$g_l(x_\lambda)=s_\lambda+(1-\lambda)s'_\lambda, \quad g_G(x_\lambda)=s_\lambda-\lambda s'_\lambda$$

$$s''_\lambda = -\frac{2(1-m_\lambda)[r+2m_\lambda(\alpha-\theta^2)]}{1+2\alpha_\lambda m_\lambda}<0.$$

Primes mean differentiation in λ at fixed α, r. Then

$$\frac{d}{d\lambda}g_l(x_\lambda)=(1-\lambda)s''_\lambda\le 0, \qquad \frac{d}{d\lambda}g_G(x_\lambda)=-\lambda s''_\lambda\ge 0.$$

Proof. From the definition of s_{NN},

$$\frac{d}{d\alpha}s_{NN}(\alpha)=m_{\alpha,NN}, \qquad \frac{d}{d\alpha}m_{\alpha,NN}=\frac{m_{\alpha,NN}(1-m_{\alpha,NN})}{1+2\alpha m_{\alpha,NN}}.$$

Therefore

$$\alpha'_\lambda=\alpha-1-(1-2\lambda)r,$$

and

$$s'_\lambda = m_\lambda \alpha'_\lambda + (1-2\lambda)r.$$

Differentiating once more gives

$$s''_\lambda = \frac{m_\lambda(1-m_\lambda)(\alpha'_\lambda)^2}{1+2\alpha_\lambda m_\lambda} - 2r(1-m_\lambda).$$

Using

$$4r\alpha_\lambda - (\alpha'_\lambda)^2 = 4(\alpha - \theta^2),$$

we obtain

$$s''_\lambda = -\frac{2(1-m_\lambda)\left[r + 2m_\lambda(\alpha-\theta^2)\right]}{1+2\alpha_\lambda m_\lambda} < 0,$$

since $r > 0$ and feasibility gives $\theta^2 \le \alpha$.

Next, direct substitution of x_λ gives

$$g_l(x_\lambda; S) - g_G(x_\lambda; S) = m_\lambda(\alpha - 1) + (1-2\lambda)(1-m_\lambda)r = s'_\lambda.$$

Together with

$$\lambda g_l(x_\lambda; S) + (1-\lambda) g_G(x_\lambda; S) = s_\lambda,$$

we solve the two linear equations and obtain

$$g_l(x_\lambda; S) = s_\lambda + (1-\lambda)s'_\lambda,$$

and

$$g_G(x_\lambda; S) = s_\lambda - \lambda s'_\lambda.$$

Finally,

$$\frac{d}{d\lambda} g_l(x_\lambda; S) = (1-\lambda)s''_\lambda \le 0,$$

while

$$\frac{d}{d\lambda} g_G(x_\lambda; S) = -\lambda s''_\lambda \ge 0.$$

Hence $g_l(x_\lambda; S)$ strictly decreases and $g_G(x_\lambda; S)$ strictly increases with λ. □

For later calculation, we will need the explicit expressions of the two distances. They are

$$g_l(x_\lambda; S) = \alpha m_\lambda^2 + m_\lambda + (1-\lambda)(1-m_\lambda)\Big[m_\lambda(\alpha + r - 1) + (1-\lambda)(1-m_\lambda)r\Big],$$
$$g_G(x_\lambda; S) = m_\lambda^2 + m_\lambda + \lambda(1-m_\lambda)\Big[m_\lambda(1 + r - \alpha) + \lambda(1-m_\lambda)r\Big].$$

The following lemma introduces $s_{\text{crit}}(\alpha, r)$ which is the minimum radius in the threshold graph required to guarantee that we can make progress towards the query given the norm of the current vector (α) and the current distance to the query (r). Notice that this threshold distance implies a corresponding value of ρ_s.

Lemma 12. *Fix a feasible state*

$$S = (\alpha, 1, \theta), \quad \alpha \in [\alpha_{min}, \alpha_{max}], \quad r = \alpha + 1 - 2\theta > s_{NN}(1).$$

Define the critical local radius

$$s_{crit}(\alpha, r) := \min\left\{ g_l(x; S) : J(x) \le \frac{1}{w} \quad g_G(x; S) \le r \right\}.$$

Then,

$$s_{crit}(\alpha, r) = \begin{cases} s_{NN}(\alpha), & r \ge 2 - \alpha(1 - m_{\alpha,NN}) \\ g_l(x_{\lambda_*}; S), & otherwise \end{cases}$$

where the $\lambda_* \in (0,1)$ is the unique solution of

$$g_G(x_{\lambda_*}; S) = r.$$

Proof. We first check the two end point of the navigate family. When $\lambda = 0$ we are in fact solving

$$\min_x g_G(x; S) \quad \text{s.t. } J(x) \leq \frac{1}{w},$$

which is give x_0, the closest nearest neighbour candidate to the query. i.e. $g_G(x_0; S) = s_{NN}(1) < r$.

When $\lambda = 1$ we are in fact solving

$$\min_x g_l(x; S) \quad \text{s.t. } J(x) \leq \frac{1}{w},$$

which is give x_1, the closest nearest neighbour candidate to the source. i.e. $g_l(x_1; S) = s_{NN}(\alpha)$
Case 1: Substitution of $x_1 = ((1 - m_1)\sqrt{\alpha}, 0, m_1)$ into the query term we have

$$g_G(x_1; S) = \alpha(1 - m_1)^2 - 2(1 - m_1)\theta + m_1 + 1.$$

Now check

$$\begin{aligned} r - g_G(x_1; S) &= \alpha + 1 - 2\theta - \left[\alpha(1 - m_1)^2 - 2(1 - m_1)\theta + m_1 + 1\right] \\ &= m_1(\alpha(1 - m_1) + r - 2) \end{aligned}$$

Then

$$r - g_G(x_1; S) \geq 0 \iff r \geq 2 - \alpha(1 - m_1).$$

Suppose such condition hole, then the greedy condition is satisfied and x_1 is feasible for such optimization. But notice that x_1 already minimizes g_l over the $J(x) \leq \frac{1}{w}$. Using $m_1 = m_{\alpha,NN}$ we have $s_{crit}(\alpha, r) = g_l(x_1; S) = s_{NN}(\alpha)$. The uniqueness follows from Lemma 4, Theorem 7.

Case 2: Now suppose

$$r < 2 - \alpha(1 - m_{\alpha,NN}).$$

Then we have

$$g_G(x_0; S) < r < g_G(x_1; S).$$

Lemma 11 gives $g_G(x_\lambda; S)$ is continuous and strictly increasing in λ. Hence there exist a unique

$$\lambda_* \in (0, 1) \quad s.t. \quad g_G(x_{\lambda_*}; S) = r.$$

By Theorem 8 we have

$$\lambda_* g_l(x; S) + (1 - \lambda_*) g_G(x; S) \geq \lambda_* g_l(x_{\lambda_*}; S) + (1 - \lambda_*) g_G(x_{\lambda_*}; S).$$

As $\lambda_* > 0$ and $g_G(x_{\lambda_*}; S) = r$,

$$\begin{aligned} \lambda_* \left[g_l(x; S) - g_l(x_{\lambda_*}; S)\right] \geq &(1 - \lambda_*) \left[g_G(x_{\lambda_*}; S) - g_G(x; S)\right] \\ = &(1 - \lambda_*) \left[r - g_G(x; S)\right] \geq 0. \end{aligned}$$

Thus we have $g_l(x; S) \geq g_l(x_{\lambda_*}; S)$ which gives

$$s_{crit}(\alpha, r) = g_l(x_{\lambda_*}; S).$$

The uniqueness of minimizer follows from Theorem 8. □

Remark 4. *At this point it is essential to summarize about what we have done in this subsection. We first fixed λ and found x_λ to solve the optimization problem*

$$\min_x [\lambda g_l(x; S) + (1 - \lambda) g_G(x; S)] \quad \textit{s.t. } J(x) \leq \frac{1}{w},$$

which in fact is the nearest neighbor distance of the imaginary C_λ. Then we picked a value of λ carefully, so that $g_G(x_{\lambda_*}; S) = r$. This ensures that we guarantee a candidate which will make progress towards the query.

In order the obtain the required degree of the graph, we will need to use the local term $g_l(x_{\lambda_*}; S)$. In the following subsection, we will introduce the critical value of ρ_s and discuss its properties.

B.2 Critical Threshold Properties and Critical Degree Exponent

The previous subsection solves the minimization problem for every fix α and r, and establishes the critical threshold distances $s_{\rm crit}(\alpha, r)$ which must be used in the threshold graph. In this subsection we will find a condition so that we can find the worst case critical threshold and establish the value of ρ_s that we must use.

Lemma 13. *For a feasible state (α, r) with $r > s_{\rm NN}(1)$, we have*

$$\frac{\partial}{\partial \alpha} s_{crit}(\alpha, r) > 0\,, \quad \frac{\partial}{\partial r} s_{crit}(\alpha, r) \leq 0.$$

Moreover, if $r < 2 - \alpha(1 - m_{\alpha,\rm NN})$, let $m_{crit} := m_\alpha(s_{crit}(\alpha, r))$, and we have

$$\frac{d}{d\alpha} I_\alpha(s_{crit}(\alpha, r)) = \frac{1 - m_{crit}}{2m_{crit}} \left(m_{crit} - m_{\lambda_*}\right).$$

Otherwise, we have

$$\frac{d}{d\alpha} I_\alpha(s_{crit}(\alpha, r)) = 0.$$

Proof. Recall the two cases in the Lemma 12. We first consider the **case 2**. It gives a unique $\lambda_* \in (0, 1)$ then

$$s_{\lambda_*} = \lambda_* s_{crit}(\alpha, r) + (1 - \lambda_*) r\,, \quad s'_{\lambda_*} = s_{crit}(\alpha, r) - r.$$

With $s_\lambda = s_{NN}(\alpha_\lambda) + \lambda(1 - \lambda)r$ we have

$$\frac{\partial}{\partial \alpha} s_\lambda = \frac{\partial s_\lambda}{\partial \alpha_\lambda} \frac{\partial \alpha_\lambda}{\partial \alpha} = \lambda m_\lambda,$$

and

$$\frac{\partial}{\partial r} s_\lambda = \lambda(1 - m_\lambda)(1 - \lambda).$$

Consider fix r first

$$\begin{aligned}
\frac{\partial}{\partial \alpha} s_{\lambda_*} &= \frac{\partial \lambda_*}{\partial \alpha} s_{crit}(\alpha, r) + \lambda_* \frac{\partial s_{crit}(\alpha, r)}{\partial \alpha} + \frac{\partial}{\partial \alpha}(1 - \lambda_*) r \\
\frac{\partial}{\partial \alpha} s_\lambda \Big|_{\lambda = \lambda_*} + \frac{\partial s_{\lambda_*}}{\partial \lambda_*} \frac{\partial \lambda_*}{\alpha} &= \frac{\partial \lambda_*}{\partial \alpha} (s_{crit}(\alpha, r) - r) + \lambda_* \frac{\partial s_{crit}(\alpha, r)}{\partial \alpha} \\
\lambda_* \frac{\partial s_{crit}(\alpha, r)}{\partial \alpha} &= \frac{\partial}{\partial \alpha} s_\lambda \Big|_{\lambda = \lambda_*} \\
\frac{\partial}{\partial \alpha} s_{crit}(\alpha, r) &= m_{\lambda_*} > 0.
\end{aligned}$$

Now fix α we have

$$\begin{aligned}
\frac{\partial}{\partial r} s_{\lambda_*} &= \frac{\partial \lambda_*}{\partial r} s_{crit}(\alpha, r) + \lambda_* \frac{\partial s_{crit}(\alpha, r)}{\partial r} + \frac{\partial}{\partial r}(1 - \lambda_*) r \\
\lambda_*(1 - m_{\lambda_*})(1 - \lambda_*) &= \lambda_* \frac{\partial s_{crit}(\alpha, r)}{\partial r} + (1 - \lambda_*) \\
\frac{\partial}{\partial r} s_{crit}(\alpha, r) &= -(1 - \lambda_*) \left(m_{\lambda_*} + \frac{1 - \lambda_*}{\lambda_*} \right) < 0
\end{aligned}$$

Now we consider $I_\alpha(s_{crit}(\alpha, r))$. Theorem 6 gives

$$2 \frac{\partial I_\alpha}{\partial \alpha} = \frac{\partial m}{\partial \alpha} - \frac{1}{m} \frac{\partial m}{\partial \alpha} + (m - 1)^2 + \alpha \cdot 2(m - 1) \frac{\partial m}{\partial \alpha},$$

with

$$\frac{\partial m}{\partial \alpha} = -\frac{m^2}{1 + 2\alpha m},$$

we have

$$\frac{\partial I_\alpha(s)}{\partial \alpha} = \frac{1 - m_\alpha(s)}{2}.$$

Notice that we fixed α in Theorem 6 but now we move it. A useful mindset is see the rate function as $I(\alpha, s)$ where s is also a function of α. We also need

$$\begin{aligned}\frac{\partial I_\alpha(s)}{\partial s} &= \frac{\partial I_\alpha(s)}{\partial m}\frac{\partial m}{\partial s} \\ &= -\frac{1-m}{2m}.\end{aligned}$$

Now we have

$$\begin{aligned}\frac{d}{d\alpha} I_\alpha(s_{crit}(\alpha, r)) &= \frac{\partial I_\alpha(s_{crit}(\alpha, r))}{\partial \alpha} + \frac{\partial I_\alpha(s_{crit}(\alpha, r))}{\partial s_{crit}(\alpha, r)}\frac{\partial s_{crit}(\alpha, r)}{\partial \alpha} \\ &= \frac{1 - m_{crit}}{2} + -\frac{1 - m_{crit}}{2m_{crit}} m_{\lambda_*} \\ &= \frac{1 - m_{crit}}{2m_{crit}} \left(m_{crit} - m_{\lambda_*}\right).\end{aligned}$$

We finally consider **Case 1**. By Lemma 12 we have

$$s_{crit}(\alpha, r) = s_{NN}(\alpha).$$

Then

$$\frac{\partial}{\partial \alpha} s_{crit}(\alpha, r) = m_{\alpha,NN} > 0\,, \quad \frac{\partial}{\partial r} s_{crit}(\alpha, r) = 0\,.$$

Moreover recall Definition 6

$$I_\alpha(s_{crit}(\alpha, r)) = I_\alpha(s_{NN}(\alpha)) = \frac{1}{w},$$

which gives

$$\frac{d}{d\alpha} I_\alpha(s_{crit}(\alpha, r)) = 0.$$

At the boundary between the two cases, $\lambda_* \uparrow 1$. Hence

$$m_{\lambda_*} \to m_{\alpha,NN}, \qquad \frac{\partial s_{crit}}{\partial r} \to 0,$$

so the derivatives from the two cases agree. □

Lemma 14. *Let w_0 be the unique positive solution that make*

$$\alpha_{min} + s_{\mathrm{NN}}(1) = 1\,.$$

Then, for every feasible state with $r < 2 - \alpha(1 - m_{\alpha,\mathrm{NN}})$, if $w \geq w_0$ we have

$$\frac{d}{d\alpha} I_\alpha(s_{crit}(\alpha, r)) > 0\,.$$

Proof. We first show $\alpha_{min} + s_{NN}(1)$ is strictly increasing in w. Recall Lemma 7, Lemma 4 and Definition 6 we know

$$I_0(\alpha_{min}) = \frac{1}{2}(\alpha_{min} - 1 - \log(\alpha_{min})) = \frac{1}{w}$$

and

$$I_1(s_{NN}(1)) = \frac{1}{2}(m_{1,NN} - 1 - \log(m_{1,NN}) + (m_{1,NN} - 1)^2) = \frac{1}{w}\,.$$

Since $I_\alpha(s)$ is strictly decreasing on s and $1/w$ is decreasing in w. Thus both α_{min} and $s_{NN}(1)$ is increasing in w. Therefore $\alpha_{min} + s_{NN}(1)$ is strictly increasing in w. Thus by definition we have

$$w \geq w_0 \implies \alpha_{min} + s_{NN}(1) \geq 1\,.$$

Now we ready to consider the feasible state in **Case 2** which gives

$$\alpha \geq \alpha_{min}\,, \quad r > s_{NN}(1)\,.$$

Recall

$$s_{crit}(\alpha, r) = \alpha m_{\lambda_*}^2 + m_{\lambda_*} + \underbrace{(1-\lambda_*)(1-m_{\lambda_*})\Big[m_{\lambda_*}(\alpha + r - 1) + (1-\lambda_*)(1-m_{\lambda_*})r\Big]}_{\text{always } >0 \text{ when } \alpha+r-1>0.\ w_0 \text{ solve this inequality.}}\,.$$

Thus we have

$$s_{crit}(\alpha, r) = \alpha m_{crit}^2 + m_{crit} > \alpha m_{\lambda_*}^2 + m_{\lambda_*}\,.$$

Finally, $\alpha m^2 + m$ is strictly increasing for $m > 0$ thus

$$m_{crit} > m_{\lambda_*}.$$

Recall Lemma 4 we know that as long as $s_{crit}(\alpha, r)$ is less than $\alpha + 1$ we can claim $m_{crit} < 1$. By Lemma 13 we consider the $r \downarrow s_{NN}(1)$ which gives $\lambda_* \downarrow 0$. At $\lambda = 0$ we have

$$g_l(x_0) = m_0^2 + r(1 - m_0) + m_0\alpha\,.$$

Check

$$\begin{aligned}\alpha + 1 - g_l(x_0) &= \alpha + 1 - \big[m_0^2 + r(1-m_0) + m_0\alpha\big] \\ &= (1-m_0)(1 + \alpha + m_0 - r) \\ &= (1-m_0)(1 + \alpha - m_0^2) > 0\end{aligned}$$

This gives $s_{crit}(\alpha, r) < \alpha + 1$ and together with $m_{crit} > m_{\lambda_*}$ we finish the proof

$$\frac{d}{d\alpha} I_\alpha(s_{crit}(\alpha, r)) = \frac{1 - m_{crit}}{2m_{crit}}\,(m_{crit} - m_{\lambda_*}) > 0$$

□

The next lemma establishes an expression for the critical value of ρ_s which will allow the greedy search to make progress at every step. The critical value depends only on w, and c.

Theorem 9. *Let $w \geq w_0,\ c > 1$. Define the target distance*

$$T := c^2 s_{NN}(1),$$

and the critical degree

$$\rho_{crit}(c, w) = \begin{cases} 0, & T \geq 2 - \alpha_{min}(1 - m_{\alpha_{min}, NN}) \\ 1 - wI_{\alpha_{min}}(s_{crit}(\alpha_{min}, T)), & \textit{otherwise}. \end{cases}$$

For any constant $\epsilon_\rho \in (0, 1 - \rho_{crit}(c, w))$, let

$$\rho_s = \rho_{crit}(c, w) + \epsilon_\rho,$$

then at every feasible non-target state $\alpha \in [\alpha_{min}, \alpha_{max}]$ and $r \geq T$ we have

$$s_*(\alpha; \rho_s) > s_{crit}(\alpha, r).$$

Proof. Let

$$\rho_{req}(\alpha, r, c, w) := 1 - wI_\alpha(s_{crit}(\alpha, r)).$$

which we will write is as $\rho_{req}(\alpha, r)$ when context is clear. We want to reduce both α and the current term r to one state by discussing its properties. Lemma 4 gives $I_\alpha(s)$ is decreasing in s, with Lemma 13 we have

$$\frac{d}{dr}\rho_{req}(\alpha, r, c, w) = -w\,\frac{\partial I_\alpha(s)}{\partial s}\bigg|_{s=s_{crit}} \frac{\partial s_{crit}}{\partial r} \leq 0.$$

By Lemma 14

$$\frac{d}{d\alpha}\rho_{req}(\alpha, r, c, w) = -w\frac{\partial I_\alpha(s_{crit})}{\partial \alpha} \leq 0.$$

$\rho_{req}(\alpha, r)$ is not increasing in both α and r. This means we we only need to inspect the minimum value of both term to find the worst case. By definition and continuity we have

$$\rho_{req}(\alpha, r) \leq \rho_{req}(\alpha_{min}, T) = 1 - wI_{\alpha_{min}}(s_{crit}(\alpha_{min}, T)) .$$

Now recall **Case 1** in Lemma 12, whenever $T \geq 2 - \alpha_{min}(1 - m_{\alpha_{min},NN})$ we have the $s_{crit}(\alpha_{min}, T) = s_{NN}(\alpha_{min})$. And Definition 6 gives

$$I_{\alpha_{min}}(s_{NN}(\alpha_{min})) = \frac{1}{w} ,$$

so

$$\rho_{req}(\alpha_{min}, T) = 0 .$$

Otherwise, in **Case 2** we have

$$\rho_{req}(\alpha_{min}, T) = 1 - wI_{\alpha_{min}}(s_{crit}(\alpha_{min}, T)) .$$

Then we just define such worst case as $\rho_{crit}(c, w)$(only c and w matter in this term!) and claim

$$\rho_s = \rho_{crit}(c, w) + \epsilon_\rho > \rho_{req}(\alpha, r) , \quad \forall \alpha \in [\alpha_{min}, \alpha_{max}],\ r > T .$$

Then at every feasible non-target state

$$s_*(\alpha; \rho_s) > s_{crit}(\alpha, r) .$$

□

In the following Lemma, we see that ρ_{crit} is a decreasing function of c. That is, if we requiring a looser approximation of the nearest neighbor of the query, then we can afford to use a lower degree graph.

Lemma 15. *Let $w \geq w_0$. Define*

$$\tilde{c}(w) := \sqrt{\frac{2 - \alpha_{\min}(1 - m_{\alpha_{\min},\text{NN}})}{s_{\text{NN}}(1)}} ,$$

and notice that $\tilde{c}(w) > 1$.

The function $\rho_{\text{crit}}(c, w)$ is continuous and non-increasing on c. More precisely,

$$\rho_{\text{crit}}(c, w) = 0 \text{ for } c \geq \tilde{c}(w) \text{ and it is strictly decreasing on } 1 < c < \tilde{c}(w) .$$

Proof. We first show $\tilde{c}(w) > 1$. A useful result is

$$\frac{d}{d\alpha}\alpha(1 - m_{\alpha,NN}) > 0 .$$

With $\alpha_{min} < 1$ and Definition 6 we have

$$2 - \alpha_{min}(1 - m_{\alpha_{min},NN}) > 1 + m_{1.NN} > m_{1,NN} + m_{1,NN}^2 = s_{NN}(1).$$

This gives $\tilde{c}(w) > 1$.

Now recall $T = c^2 s_{NN}(1)$. If $T = c^2 s_{NN}(1) \geq 2 - \alpha_{min}(1 - m_{\alpha_{min},NN})$ we have $\rho_{crit} = 0$ by definition. The condition is equivalent to $c \geq \tilde{c}(w)$.

Now we consider $1 < c < \tilde{c}(w)$ which is exactly the **Case 2** in Theorem 9. Take derivatives we have

$$\frac{d}{dc}\rho_{crit}(c, w) = -w\frac{\partial I_{\alpha_{min}}}{\partial s}\frac{\partial s_{crit}}{\partial r}\frac{\partial T}{\partial c}.$$

Lemma 13 and auxiliary result gives

$$\frac{\partial I_{\alpha_{min}}}{\partial s} < 0\,,\ \frac{\partial s_{crit}}{\partial r} < 0\,,\ \frac{\partial T}{\partial c} = 2c \cdot s_{NN}(1) > 0\,.$$

This gives

$$\frac{d}{dc}\rho_{crit}(c,w) < 0$$

which prove the strictly decreasing.

When $c \uparrow \tilde{c}(w)$ we have $T \uparrow 2 - \alpha_{min}(1 - m_{\alpha_{min},NN})$, then two case in Theorem 9 meet at

$$s_{crit}(\alpha_{min}, 2 - \alpha_{min}(1 - m_{\alpha_{min},NN})) = s_{NN}(\alpha_{min}).$$

Hence

$$\rho_{crit} \to 1 - wI_{\alpha_{min}}(s_{NN}(\alpha_{min})) = 0.$$

This prove the continuity at $\tilde{c}(w)$. □

In the following two Lemmas, we can see that even as c approaches 1 (which would be exact nearest-neighbor recovery), the value of ρ_{crit} approaches a constant less than 1, which depends on w.

Lemma 16. *Let $w \geq w_0$, then*

$$\lim_{c\downarrow 1} \rho_{crit}(c,w) \in (0,1)$$

exists and

$$\lim_{c\downarrow 1} \rho_{\text{crit}}(c,w) = 1 - wI_{\alpha_{\text{min}}}(s_0(w))$$

where

$$s_0(w) := m_{1,\text{NN}}^2 + (1 - m_{1,\text{NN}})s_{\text{NN}}(1) + \alpha_{\text{min}} m_{1,\text{NN}}\,.$$

Proof. As $c \downarrow 1$ gives $T \downarrow s_{NN}(1)$, Lemma 12 Theorem 9 gives $\lambda_* \downarrow 0$. By continuity

$$s_{crit}(\alpha_{min}, T) = g_l(x_{\lambda_*}) \to g_l(x_0).$$

At $\lambda = 0$ we have

$$g_l(x_0) = m_{1,NN}^2 + (1 - m_{1,NN})s_{NN}(1) + \alpha_{min} m_{1,NN} = s_0(w)\,.$$

By continuity we have

$$\lim_{c\downarrow 1} \rho_{crit}(c,w) = 1 - wI_{\alpha_{min}}(s_0(w))\,.$$

What remain to check is its range. Recall the auxiliary result in Lemma 14 we have

$$s_0(w) < 1 + \alpha_{min}\,,$$

which gives $I_{\alpha_{min}}(s_0(w)) < \frac{1}{w}$. Thus $1 - wI_{\alpha_{min}}(s_0(w)) < 1$.

On the other hand, $w \geq w_0$ gives $\alpha_{min} + s_{NN}(1) > 1$.Then use the explicit expression of $g_l(x_0)$,

$$\begin{aligned} s_0(w) &= \alpha_{min} m_{1,NN}^2 + m_{1,NN} \\ &\quad + (1 - m_{1,NN})\Big[m_{1,NN}(\alpha_{min} + s_{NN}(1) - 1) + (1 - m_{1,NN})s_{NN}(1)\Big] \\ &> \alpha_{min} m_{1,NN}^2 + m_{1,NN}. \end{aligned}$$

Since $\alpha_{min} < 1$ and$m_{\alpha,NN}$ is strictly increasing in α,

$$m_{\alpha_{min},NN} < m_{1,NN}.$$

As $m \mapsto \alpha_{min} m^2 + m$ is strictly increasing,

$$\alpha_{min} m_{1,NN}^2 + m_{1,NN} > s_{NN}(\alpha_{min}).$$

Hence

$$s_0(w) > s_{NN}(\alpha_{min}).$$

Since $I_{\alpha_{min}}$ is strictly decreasing,

$$I_{\alpha_{min}}(s_0(w)) < I_{\alpha_{min}}(s_{NN}(\alpha_{min})) = \frac{1}{w}.$$

Therefore

$$\rho_{exact}(w) > 0.$$

□

Lemma 17. *We have*

$$\lim_{c\downarrow 1} \rho_{\mathrm{crit}}(c, w) = \frac{8}{9} + O(w^{-1/2}).$$

Moreover,

$$\tilde{c}(w) = 1 + \frac{1}{\sqrt{3w}} + O(w^{-1}).$$

Consequently, for every fixed $c > 1$, there exists a constant w such that

$$\rho_{\mathrm{crit}}(c, w) = 0.$$

Proof. We first record the large-w behaviour of the basic quantities. From

$$\frac{1}{2}(\alpha_{min} - 1 - \log \alpha_{min}) = \frac{1}{w}$$

and the Taylor expansion $-\log(1-x) = x + \frac{x^2}{2} + O(x^3)$, we obtain

$$\alpha_{min} = 1 - \frac{2}{\sqrt{w}} + O(w^{-1}).$$

Similarly, from the defining equation of $m_{1,NN}$,

$$m_{1,NN} = 1 - \frac{2}{\sqrt{3w}} + O(w^{-1}),$$

and the same calculation gives

$$m_{\alpha_{min},NN} = 1 - \frac{2}{\sqrt{3w}} + O(w^{-1}).$$

Therefore

$$s_{NN}(1) = m_{1,NN}^2 + m_{1,NN} = 2 - \frac{2\sqrt{3}}{\sqrt{w}} + O(w^{-1}),$$

while

$$2 - \alpha_{min}(1 - m_{\alpha_{min},NN}) = 2 - \frac{2}{\sqrt{3w}} + O(w^{-1}).$$

Hence

$$\begin{aligned}\tilde{c}(w) &= \sqrt{\frac{2 - \frac{2}{\sqrt{3w}} + O(w^{-1})}{2 - \frac{2\sqrt{3}}{\sqrt{w}} + O(w^{-1})}} \\ &= 1 + \frac{1}{\sqrt{3w}} + O(w^{-1}).\end{aligned}$$

We next consider $c \downarrow 1$. By Lemma 16,

$$\lim_{c\downarrow 1} \rho_{crit}(c, w) = 1 - wI_{\alpha_{min}}(s_0(w)),$$

where

$$s_0(w) = m_0^2 + (1 - m_0)s_{NN}(1) + \alpha_{min} m_0, \qquad m_0 = m_{1,NN}.$$

Let

$$\Delta := 1 + \alpha_{min} - s_0(w).$$

Then

$$\Delta = (1 - m_0)(1 + \alpha_{min} - m_0^2) = \frac{2}{\sqrt{3w}} + O(w^{-1}).$$

For $s = \alpha + 1 - \Delta$, the rate function has the expansion

$$I_\alpha(s) = \frac{\Delta^2}{4(1 + 2\alpha)} + O(\Delta^3).$$

Thus

$$I_{\alpha_{min}}(s_0(w)) = \frac{1}{9w} + O(w^{-3/2}),$$

and consequently

$$\lim_{c\downarrow 1} \rho_{crit}(c, w) = 1 - \frac{1}{9} + O(w^{-1/2}) = \frac{8}{9} + O(w^{-1/2}).$$

Finally,

$$\tilde{c}(w) \to 1.$$

Hence for every fixed $c > 1$, we have $c > \tilde{c}(w)$ for all sufficiently large w. By Lemma 15,

$$\rho_{crit}(c, w) = 0.$$

□

C From Navigate family to Greedy Search

In the previous section we solved the one-step progression problem in continuous space. However, the optimal threshold distance s_{crit} may only guarantee that $g_G(x) = r$. In other words, the new distance to the query is equal the current distance. In this section, we will first extend this continuous result to show that there is a candidate which makes strict progress towards the query. Then we will transfer the continuous argument to show that there indeed is a corresponding candidate in the sampled data set with high probability. Finally, we transfer the result on the thresholCd graph back to the KNN graph.

C.1 Constant Step Greedy Search

We begin by showing that there is a candidate in the continuous model which makes a strict improvement in the distance to the query.

Lemma 18. *Let $w \geq w_0$, $c > 1$ and ρ_s be defined as in Theorem 9. Then for every feasible non-target state there exists candidate $\hat{x}$ satisfying*

$$J(\hat{x}) < \frac{1}{w}, \quad g_l(\hat{x}) < s_*(\alpha; \rho_s), \quad g_G(\hat{x}) < r\,.$$

Proof. By Theorem 9 we have

$$s_*(\alpha; \rho_s) > s_{crit}(\alpha, r).$$

Let's consider the **Case 2** first. Recall that there exist a unique $\lambda_* \in (0, 1)$ so that

$$g_G(x_{\lambda_*}) = r, \quad g_l(x_{\lambda_*}) = s_{crit}(\alpha, r), \quad J(x_{\lambda_*}) = \frac{1}{w}.$$

Now, by Lemma 11 we know

$$\frac{d}{d\lambda} g_l(x_\lambda) < 0, \qquad \frac{d}{d\lambda} g_G(x_\lambda) > 0.$$

Hence we choose a $\lambda_- < \lambda_*$ that sufficiently close to λ_*, then

$$g_G(x_{\lambda_-}) < r\,,$$

while

$$s_{crit}(\alpha, r) = g_l(x_{\lambda_*}) < g_l(x_{\lambda_-}) < s_*(\alpha; \rho_s).$$

And by definition we still have

$$J(x_{\lambda_-}) = \frac{1}{w}.$$

In **Case 1**,

$$s_{crit}(\alpha, r) = s_{NN}(\alpha),$$

then

$$g_G(x_1) = r, \quad g_l(x_1) \le s_{NN}(\alpha), \quad J(x_1) = \frac{1}{w}.$$

We just apply the same trick to obtain λ_-. It remains to make the rate inequality strick. We use $J(0,0,1) = 0$. For some $\xi > 0$ let

$$\hat{x} := (1-\xi)x_{\lambda_-} + \xi(0,0,1),$$

then by convexity,

$$J(\hat{x}) \le (1-\xi)J(x_{\lambda_-}) + \xi J(0,0,1) = \frac{1-\xi}{w} < \frac{1}{w}.$$

Since both g_l, g_G are continuous and x_{λ_-} create strict inequalities, for sufficiently small $\xi > 0$ we still have

$$g_l(\hat{x}) < s_*(\alpha; \rho_s)\,, \quad g_G(\hat{x}) < r.$$

□

The following lemma establishes that we can reduce the distance to the query by a constant amount at every step of the search.

Lemma 19. *With the same assumptions as Lemma 18, there exist constants*

$$h > 0\,, \quad \eta > 0\,,$$

such that the following claim holds.

If

$$\left| \frac{\|q\|^2}{d} - 1 \right| \le h\,,$$

then, with probability $1 - o(1)$*, for every data point* X_i *that satisfies*

$$\frac{\|X_i\|^2}{d} \in [\alpha_{min} - h, \alpha_{max} + h]\,, \quad \frac{\|X_i - q\|^2}{d} > T\,,$$

there exists an out neighbor X_j *such that*

$$\frac{\|X_j - q\|^2}{d} \le \frac{\|X_i - q\|^2}{d} - \eta\,.$$

Proof. We first consider $\beta = 1$ (i.e. $h = 0$) then fix it by continuity later. Define the set of feasible state

$$\mathcal{D} := \left\{(\alpha, r) : \alpha \in [\alpha_{min}, \alpha_{max}],\ T \le r \le (\sqrt{\alpha_{max}} + 1)^2, (\alpha + 1 - r)^2 \le 4\alpha\right\},$$

which is a compact set.

Lemma 18 for every $(\alpha, r) \in \mathcal{D}$ there exist a $\hat{x}$ satisfying

$$J(\hat{x}) < \frac{1}{w}, \quad g_l(\hat{x}) < s_*(\alpha; \rho_s), \quad g_G(\hat{x}) < r\,.$$

Since those inequalities are strict, by continuity, around every state (α, r) we can choose a state neighbourhood set $\mathcal{V}$ and an open candidate set $\mathcal{B}$ such that for every state $(\tilde{\alpha}, \tilde{r}) \in \mathcal{V}$ and every $\tilde{x} \in \mathcal{B}$ satisfy

$$g_l(\tilde{x}) < s_*(\tilde{\alpha}; \rho_s)\,, \quad g_G(\tilde{x}) < \tilde{r} - \eta_{\mathcal{V}}$$

for some $\eta_{\mathcal{V}} > 0$. Besides, every $\tilde{x} \in \mathcal{B}$ so that

$$J(\tilde{x}) < \frac{1}{w} - \delta_{\mathcal{V}}$$

for some $\delta_{\mathcal{V}} > 0$.

Since $\mathcal{D}$ is compact, finitely many such state neighourhood sets $\mathcal{V}_1, \cdots, \mathcal{V}_L$ cover $\mathcal{D}$. Let the corresponding candidate set be $\mathcal{B}_1, \cdots, \mathcal{B}_L$.

Since the collection is finite, we then define

$$\eta := \min_{l \le L} \eta_{\mathcal{V}_l} > 0\,, \quad \delta := \min_{l \le L} \delta_{\mathcal{V}_l} > 0.$$

Again, since g_l, g_G and $s_*(\alpha;\rho_s)$ are continuous, slack with a sufficient small $h > 0$, the same finite collection of sets still covers every feasible state with

$$\alpha \in [\alpha_{min} - h, \alpha_{max} + h]\,, \quad |\beta - 1| \le h\,, \quad r \ge T.$$

For such state, one of the candidate set $\mathcal{B}_l$ satisfies, for every $x \in \mathcal{B}_l$ we have

$$J(x) < \frac{1}{w} - \delta, \quad g_l(x) < s_*(\alpha;\rho_s), \quad g_G(x) < r - \eta\,.$$

Now we only need to show that one of the actual sampled data points falls into the required candidate set.

For a arbitrary source point X_i and condition on X_i and q. For every remaining data point X_j ($j \neq i$), the corresponding random candidate (A_d, B_d, M_d) satisfies the LDP with rate function J as in Theorem 5.

For current source point, let $\mathcal{B}_o$ be the corresponding candidate set. If there is any sampled candidate in $\mathcal{B}_o$ then we good. So we need to consider the probability of such candidate exist.

Define

$$p_o := \mathbb{P}((A_d, B_d, M_d) \in \mathcal{B}_o).$$

Since $\mathcal{B}_o$ is open, the LDP lower bound gives

$$\liminf_{d\to\infty} \frac{1}{d} \log(p_o) \ge - \inf_{x \in \mathcal{B}_o} J(x).$$

Then by construction

$$\inf_{x \in \mathcal{B}_o} J(x) \le \frac{1}{w} - \delta\,.$$

Recall

$$\log(n) = \frac{d}{w} + o(d),$$

we have

$$(n-1)p_o \ge \exp\left(\frac{d}{w} - d\left(\frac{1}{w} - \delta\right) + o(d)\right) = \exp(\delta d + o(d))\,.$$

Then we say there exist a constant $\kappa > 0$ such that

$$(n-1)p_o \ge n^{\kappa}$$

for all sufficiently large n.

Conditional on X_i and q, we still remain $n-1$ independent data points. Hence the probability that none of them falls into $\mathcal{B}_o$ is

$$(1-p_o)^{n-1} \le \exp(-(n-1)p_o) \le \exp(-n^{\kappa})\,.$$

Such probability bound is uniform over all candidate sets $\mathcal{B}_1, \cdots, \mathcal{B}_L$. Therefore, there are at most n possible source points. Then a union bound gives

$$\mathbb{P}(\exists X_i : \forall j \neq i, X_j \notin \mathcal{B}_o) \le n \exp(-n^{\kappa}) \to 0\,.$$

This finish the claim. □

Since we make constant progress at every step, it is easy to show that the greedy search completes in $O(1)$ steps.

Theorem 10. *Fix* $w \geq w_0\,, \quad c > 1\,,$ *and*

$$\rho_s = \rho_{crit}(c, w) + \epsilon_\rho < 1$$

for some constant ϵ_ρ*. Let* q *be a query point satisfying*

$$\frac{\|q\|^2}{d} = 1 + o(1)\,.$$

Then with probability $1 - o(1)$*, greedy search on* $G_\tau(\rho_s)$ *starting from any vertex in* $\mathcal{X}$*, reaches a vertex* X_i *satisfying*

$$\frac{\|X_i - q\|^2}{d} \leq c^2 s_{NN}(1)$$

in $O(1)$ *steps. In particular, greedy search finishes the c-ANN search of* q *simultaneously for every starting vertex in* $\mathcal{X}$*.*

Proof. Let $T = c^2 s_{NN}(1)$. Then by Lemma 7 we can fix a sufficiently small $h > 0$, then with probability tends to one,

$$\forall i \leq n\,, \frac{\|X_i\|^2}{d} \in [\alpha_{min} - h,\ \alpha_{max} + h]\,.$$

And

$$\frac{\|q\|^2}{d} = 1 + o(1)$$

thus we have $|\beta - 1| \leq h$.

Then we use the high probability event of Lemma 19, whenever the current vertex has current term $r > t$, there exists an out neighbour whose gives the query term at most

$$r - \eta\,.$$

Greedy search chooses the out neighobur with smallest query term, so its next state is at least as good as droping η. We then get the simple upper bound of current term

$$r \leq \left(\sqrt{\alpha_{max} + h} + \sqrt{1 + h}\right)^2\,.$$

Hence the nuber of non-target greedy steps is at most

$$\left\lceil \frac{\left(\sqrt{\alpha_{max} + h} + \sqrt{1 + h}\right)^2 - T}{\eta} \right\rceil\,,$$

which is a constant that independent of n and d. Thus greedy search reaches the target region in $O(1)$ steps with probability tend to one. □

C.2 Back to KNN Graph

In this section we show that the greedy search on the τ-threshold graph is a sufficiently good approximation of the k-NN graph. In particular, the greedy search on the k-NN graph will also complete in $O(1)$ steps.

Lemma 20. *Consider the directed KNN graph with*

$$k = n^{\rho_k + o(1)}\,,$$

where $\rho_k > \rho_{crit}(c, w)$*. Let* ρ_t *satisfy*

$$\rho_{crit}(c, w) < \rho_t < \rho_k\,.$$

Then with probability $1 - o(1)$*, every edge in the* τ*-threshold graph with degree exponent* ρ_t *is also an edge in the KNN graph. Consequently, the greedy search on the KNN graph makes at least the same progression as the greedy search on the threshold graph.*

Proof. Fix the ρ_t and build a τ-threshold graph with it. Lemma 5 gives, for the threshold graph

$$\max_{i\le n} D_{out}(X_i) \le n^{\rho_t+o(1)}\,.$$

By definition we have $n^{\rho_t+o(1)} < n^{\rho_k+o(1)} = k$. Hence, for every source X_i, there are fewer than k data points inside its threshold. With probability tends to one,

$$E(G_\tau) \subseteq E(G_{KNN})\,.$$

Lemma 19 shows for $\rho_t > \rho_{crit}(c,w)$

$$\forall i \le n\ \exists j \ne i\ ,\ s.t.\ \frac{\|X_j - q\|^2}{d} \le \frac{\|X_i - q\|^2}{d} - \eta\,.$$

As such $X_i \to X_j$ that gives progression is also in $E(G_{KNN})$, the greedy search on KNN graph makes at least the same progression as searching on threshold graph. □

D Geometric Hubs and Navigation

In this section we will first define the geometric Hubs as a natural inherent property of KNN graphs on high-dimensional data whose distribution gives varied norms. Throughout this section, we will only consider the continuous space for convenience. In this section we assume

$$w \ge w_0\,,\quad 1 > \rho_s > 0$$

are fixed. The main objective in this section is to show that there is a difference between geometric hubs and navigating region (norm of vertices that are popular when searching). Hence we first define the geometric hub and show its behavior. Then, we consider the navigable behavior and show its relation with the geometric hub.

D.1 Geometric Hubs

In this section, we need to change our perspective compared with the previous analysis. We will still use a fix v with

$$\alpha = \frac{\|v\|^2}{d}\,,$$

but now we will consider the probability that a random point U has v as an outgoing neighbor:

$$\mathbb{P}(U \to v)$$

where the $U \sim N(0, I_d)$. Then we can calculate the in degree for v with $(n-1)\mathbb{P}(U \to v)$. Define

$$\gamma := \frac{\|U\|^2}{d}\,,\ \text{then } U \to v \iff \frac{\|U - v\|^2}{d} \le s_*(\gamma;\rho_s).$$

Lemma 21. *Let*

$$Y_d := \frac{1}{d}\sum_{j=2}^{d}\langle U, e_j\rangle^2\,.$$

Then the law of (A_d, Y_d) satisfies an LDP on $\mathbb{R}^2$ with speed d and good rate function

$$\Phi(a,y) := \begin{cases} \frac{1}{2}(a^2 + y - 1 - \log(y)), & y > 0 \\ +\infty, & y \le 0\,. \end{cases}$$

If we limit the domain to the space which would result in an incoming edge to a vertex with norm α, the rate function is

$$I_{\text{in}}(\alpha) := \inf_{a\in\mathbb{R}, y>0}\left\{\Phi(a,y) : (a - \sqrt{\alpha})^2 + y \le s_*(a^2 + y;\rho_s)\right\}.$$

This gives rise to a critical rate $I_{\text{in}}(\alpha) = \frac{1-\rho_s}{w}$ for in-degree, such that

$$\begin{cases} I_{\text{in}}(\alpha) < \frac{1-\rho_s}{w}, & \textit{we expect larger in-degree than out-degree}, \\ I_{\text{in}}(\alpha) = \frac{1-\rho_s}{w}, & \textit{we expect the same in- and out-degree}, \\ I_{\text{in}}(\alpha) > \frac{1-\rho_s}{w}, & \textit{we expect smaller in-degree than out-degree}. \end{cases}$$

Proof. We use the same proof technique as Theorem 5 to find the rate function

$$\Phi(a,y) := \begin{cases} \frac{1}{2}(a^2+y-1-\log(y)), & y>0 \\ +\infty, & y \le 0\,. \end{cases}$$

We then have

$$\gamma = a^2+y \text{ then } U \to v \iff \frac{\|U-v\|^2}{d} = (a-\sqrt{\alpha})^2 + y \le s_*(\gamma;\rho_s).$$

This give the minimum rate required to create on incoming edge to v is

$$I_{in}(\alpha) := \inf_{a\in\mathbb{R},y>0} \left\{\Phi(a,y) : (a-\sqrt{\alpha})^2 + y \le s_*(a^2+y;\rho_s)\right\}$$

Now we can use LDP to claim

$$\mathbb{P}(U \to v) = \exp\left(-dI_{in}(\alpha) + o(d)\right)\,.$$

Thus we can obtain the in-degree for v

$$\mathbb{E}[D_{in}(v)|v] = (n-1)\mathbb{P}(U\to v) = n^{1-wI_{in}(\alpha)+o(1)}\,.$$

On the other hand Lemma 5 gives

$$\mathbb{E}[D_{out}(v)|v] = n^{\rho_s+o(1)}\,.$$

Thus the critical rate is

$$\frac{\mathbb{E}[D_{in}(v)|v]}{\mathbb{E}[D_{out}(v)|v]} = 1 \implies I_{in}(\alpha) = \frac{1-\rho_s}{w}\,.$$

□

The previous lemma expresses the expected in-degree in terms of $I_{\text{in}}(\alpha)$. A smaller value of the incoming-edge rate function gives a larger expected in-degree exponent. We next show how this rate changes with the normalized squared norm of the destination vertex.

Lemma 22. *The function $I_{\text{in}}(\alpha)$ is continuous and non-decreasing in $\alpha > 0$. Moreover, whenever $I_{\text{in}}(\alpha) > 0$ we have*

$$\frac{d}{d\alpha} I_{\text{in}}(\alpha) > 0\,.$$

Proof. Recall the auxiliary results

$$\frac{d}{d\gamma} s_*(\gamma) = m_\gamma(s_*(\gamma))\,, \quad m_\gamma(s_*(\gamma)) \in (0,1)\,.$$

Let write the edge constraint as a function for convenience

$$\Psi(a,y) := (a-\sqrt{\alpha})^2 + y - s_*(a^2+y;\rho_s)\,.$$

Then

$$I_{in}(\alpha) = \inf_{a\in\mathbb{R}} \left\{\Phi(a,y) : \Psi(a,y) \le 0, y>0\right\}\,.$$

We first consider $I_{in}(\alpha) > 0$. In order to find such minimizer we can use (almost) same technique in Theorem 6 with the lagrange multiplier $\lambda_{in} > 0$ which gives

$$a + \lambda_{in}[2(a-\sqrt{\alpha}) - 2am_\gamma(s_*(\gamma))] = 0\,, \quad \frac{1}{2}(1-\frac{1}{y}) + \lambda_{in}(1-m_\gamma(s_*(\gamma))) = 0\,.$$

Write $m_\gamma = m_\gamma(s_*(\gamma))$ rearrange them

$$a[1+2\lambda_{in}(1-m_\gamma)] = 2\lambda_{in}\sqrt{\alpha}\,, \quad y = \frac{1}{1+2\lambda_{in}(1-m_\gamma)}.$$

By the evelope theorem we have

$$\frac{d}{d\alpha} I_{in}(\alpha) = \lambda_{in} \frac{\partial \Psi(a, y)}{\partial \alpha} = \lambda_{in} \left(1 - \frac{a}{\sqrt{\alpha}}\right).$$

To show it is positive we need to show $0 < a < \sqrt{\alpha}$. It is clear that $a > 0$, we can prove $a < \sqrt{\alpha}$ by contradictions. Suppose $a \geq \sqrt{\alpha}$, the equations from Lagrange Multiplier give

$$\begin{aligned} a[1 + 2\lambda_{in}(1 - m_\gamma)] &= 2\lambda_{in}\sqrt{\alpha} \\ 1 + 2\lambda_{in}(1 - m_\gamma) &\leq 2\lambda_{in} \\ 2\lambda_{in} m_\gamma &> 1\,. \end{aligned}$$

Put back to the equation of y gives $y \leq m_\gamma$. Rearrange the original equation gives

$$a - \sqrt{\alpha} = \frac{a(2\lambda_{in} m_\gamma - 1)}{2\lambda_{in}} = a(m_\gamma - \frac{1}{2\lambda_{in}}) < a m_\gamma\,.$$

Now we have

$$(a - \sqrt{\alpha})^2 + y < a^2 m_\gamma^2 + y < a^2 m_\gamma^2 + m_\gamma < (a^2 + y) m_\gamma^2 + m_\gamma\,.(*)$$

Notice that

$$m_\gamma = m_\gamma(s_*(\gamma))\,, s_*(\gamma) = \gamma m_\gamma(s_*(\gamma))^2 + m_\gamma(s_*(\gamma))\,, \gamma = a^2 + y\,.$$

Thus

$$(*) \implies (a - \sqrt{\alpha})^2 + y < s_*(a^2 + y; \rho_s),$$

which contradicts to the edge constraint! This gives

$$\frac{d}{d\alpha} I_{in}(\alpha) > 0\,.$$

Now we only need to discuss the $I_{in}(\alpha) = 0$. In this case we have $(0, 1)$ as the unique minimizer for Φ. And we have $I_{in}(\alpha) = 0$ when $(0, 1)$ satisfies the edge constraint

$$\alpha + 1 \leq s_*(1; \rho_s).$$

Combining both case show $I_{in}(\alpha)$ is non-decreasing on $(0, \infty)$. The continuity follows from the definition of I_{in} . □

To identify the geometric hub region, we need to compare $I_{\text{in}}(\alpha)$ with the edge rate $(1 - \rho_s)/w$. The following identity is obtained by computing the rate of a directed edge between two independent Gaussian points. We will first condition on its destination and then on its source.

Lemma 23. *We have*

$$\inf_{\alpha > 0} \{I_0(\alpha) + I_{\text{in}}(\alpha)\} = \frac{1 - \rho_s}{w}\,.$$

The unique minimizer is

$$\alpha_{\text{flip}} = (1 - m_{\text{flip}})^2 + m_{\text{flip}}\,, \quad m_{\text{flip}} := m_1(s_*(1; \rho_s))\,.$$

And $\alpha_{\text{flip}} < 1$.

Proof. Here we need to step back and recap the meaning of the $I_{in}(\alpha)$, it means given a current point v with normalised norm α the rate that other random point U connect v. Then $I_0(\alpha)$ means the rate of such α. In order to consider the joint probability we can add them together to claim the rate of there is a edge $U \to v$ for some v

$$I_0(\alpha) + I_{in}(\alpha)\,.$$

Then if we forget the fixed v we then LDP gives the rate of on random directed edge $U \to v$

$$\inf_{\alpha > 0} \{I_0(\alpha) + I_{in}(\alpha)\}\,.$$

Now we **flip** our mind and think such event from U point of view. In this way $U \to v$ means

$$\frac{\|v - U\|^2}{d} \leq s_*(\gamma; \rho_s)\,.$$

By definition we have the rate of it

$$I_\gamma(s_*(\gamma; \rho_s)) = \frac{1 - \rho_s}{w}\,.$$

And again we have the rate of such γ is $I_0(\gamma)$. Again we can obtain the rate (again, think of joint probability) of given U the rate of there is a edge $U \to v$ is

$$I_0(\gamma) + \frac{1 - \rho_s}{w}\,.$$

By LDP again we have the rate of on random directed edge $U \to v$

$$\inf_{\gamma > 0} I_0(\gamma) + \frac{1 - \rho_s}{w}.$$

This is exactly the same event we consider for the v, and we solve $\inf_{\gamma>0} I_0(\gamma) = 0$ when $\gamma = 1$ with same technique in Theorem 6 with the rate function in Lemma 6. And we have $m_{\text{flip}} \in (0, 1)$

$$\alpha_{\text{flip}} = (1 - m_{\text{flip}})^2 + m_{\text{flip}} = 1 - m_{\text{flip}}(1 - m_{\text{flip}}) < 1$$

This prove the claim. □

We can now identify the norm levels with larger expected in-degree. The next lemma shows that there is a unique boundary below the typical normalized squared norm, which is one.

Lemma 24. *There exists a unique $\alpha_h \in (\alpha_{min}, 1)$ such that*

$$I_{in}(\alpha_h) = \frac{1 - \rho_s}{w}\,.$$

We define the geometric hubs region

$$\mathcal{H}_{geo} = [\alpha_{min}, \alpha_h)\,.$$

Proof. The uniqueness is directly from Lemma 22. And Lemma 23 gives

$$I_0(\alpha_{flip}) + I_{in}(\alpha_{flip}) = \frac{1 - \rho_s}{w}\,.$$

Since $\alpha_{flip} < 1$ we have $I_0(\alpha_{flip}) > 0$, this gives

$$I_{in}(\alpha_{flip}) < \frac{1 - \rho_s}{w}\,.$$

Indeed

$$I_0(\alpha_{flip}) < \frac{1 - \rho_s}{w} < \frac{1}{w} = I_0(\alpha_{min})$$

gives $\alpha_{min} < \alpha_{flip} < 1$. Then

$$I_{in}(\alpha_{min}) \leq I_{in}(\alpha_{flip}) < \frac{1 - \rho_s}{w}\,.$$

As the α_{flip} is the unique minimizer, with $\alpha_{flip} \neq 1$ and $I_0(1) = 0$,

$$I_0(1) + I_{in}(1) > \frac{1 - \rho_s}{w} \implies I_{in}(1) > \frac{1 - \rho_s}{w}\,.$$

Now we have

$$I_{in}(\alpha_{min}) < \frac{1 - \rho_s}{w} < I_{in}(1)$$

by Lemma 23 we finish the proof. □

Definition 7. *We define the geometric hub region to be*

$$\mathcal{H}_{geo} := [\alpha_{min}, \alpha_h).$$

By Lemmas 21 and 22, these are exactly the norm levels in the dataset norm interval whose expected in-degree exponent is larger than ρ_s. At α_h, the expected incoming and outgoing degrees have the same leading exponent. Now the question is how large and how fast the incoming degree grows which is explain in the following proposition.

Lemma 25. *Define the geometric hub enhancement exponent*

$$\Delta(\alpha) := \rho_{in}(\alpha) - \rho_s = w\left[\frac{1-\rho_s}{w} - I_{in}(\alpha)\right].$$

Then we have

$$\Delta(\alpha_h) = 0\,,$$

and it is strictly decreasing in α wherever $I_{in}(\alpha) > 0$.

Moreover,

$$\Delta(\alpha) = wI'_{in}(\alpha_h)(\alpha_h - \alpha) + o(\alpha_h - \alpha)$$

as $\alpha \uparrow \alpha_h$.

Then we have

$$\frac{\mathbb{E}[D_{in}(v)|v]}{\mathbb{E}[D_{out}(v)|v]} = n^{\Delta(\alpha)+o(1)}\,.$$

Proof. The monotonically and endpoint is immediate from Lemma 22 and Lemma 24. We then consider

$$\Delta'(\alpha) = -wI'_{min}(\alpha) < 0$$

when $I_{in}(\alpha) > 0$. Since $I_{in}(\alpha_h) = \frac{1-\rho_s}{w} > 0$ we have $I'_{in}(\alpha_h) > 0$. Taylor expand around α_h we have

$$\begin{aligned} I_{in}(\alpha) &= I_{in}(\alpha_h) + I'_{in}(\alpha_h)(\alpha - \alpha_h) + o(|\alpha - \alpha_h|) \\ &= \frac{1-\rho_s}{w} + I'_{in}(\alpha_h)(\alpha - \alpha_h) + o(|\alpha - \alpha_h|)\,. \end{aligned}$$

Hence we have $\Delta(\alpha) = wI'_{in}(\alpha_h)(\alpha_h - \alpha) + o(\alpha_h - \alpha)$ when $\alpha \uparrow \alpha_h$. Finally Lemma 21 gives the expect degree difference. □

Remark 5. *Lemma 25 explain why the hubness phenomenon seems very extreme in both side. If we walk in to the geometric hub region for a constant drop $\epsilon > 0$ so that*

$$\alpha = \alpha_h - \epsilon.$$

This give a enhance exponent

$$\Delta(\alpha) \geq wI'(\alpha_h) \cdot \epsilon > 0\,.$$

Therefore

$$\frac{\mathbb{E}D_{in}}{\mathbb{E}D_{out}} = n^{\Delta(\alpha)+o(1)} = \exp\left(\frac{\Delta(\alpha)}{w}d + o(d)\right).$$

Vice versa.

D.2 Navigable Hubs

A large expected in-degree does not by itself determine which vertices are selected during search. We now return to a fixed source v and a query q with $\|q\|^2/d = 1$, and study the norm selected by one step of the continuous greedy model. Recall

$$\alpha = \frac{\|v\|^2}{d}\,, \quad r = \frac{\|v-q\|^2}{d}\,, \quad \theta = \frac{\alpha + 1 - r}{2}\,.$$

and write $S = (\alpha, 1, \theta)$. The model selects a candidate that minimizes the query distance subject to the probability and local constraints:

$$x_+(\alpha, r) := \arg\min \left\{ g_G(x) : J(x) \le \frac{1}{w}, \quad g_l(x) \le s_*(\alpha; \rho_s) \right\}.$$

Writing $x_+ = (a_+, b_+, m_+)$, we define its normalized squared norm by

$$\alpha_+(\alpha, r) := a_+^2 + b_+^2 + m_+.$$

We first express this next norm using the navigate family.

Lemma 26. *Let (α, r) be the feasible state, and let $x_\lambda,\ \lambda \in (0,1)$ be the navigate family from Theorem 8. Define*

$$\lambda_+ := \inf\{\lambda \in [0,1] : g_l(x_\lambda) \le s_*(\alpha; \rho_s)\}.$$

If $\lambda_+ = 0$, then $x_+ = x_0$ and α_+ is independent of r. Otherwise, we have $\lambda_+ \in (0,1)$ and

$$g_l(x_{\lambda_+}) = s_*(\alpha; \rho_s)\,.$$

In this case we have

$$\frac{\partial \alpha_+}{\partial r} - \frac{(1-\lambda)(1-m_\lambda)(1-\theta)}{r + 2m_\lambda(\alpha - \theta^2)} < 0\,.$$

Proof. We first discuss two end case in Theorem 8. At $\lambda = 1$, x_1 gives the nearest neighbour scale for current state

$$g_l(x_1) = s_{NN}(\alpha)$$

which indeed less than $s_*(\alpha; \rho_s)$ by definition. Thus λ_+ is nonempty. If $\lambda_+ = 0$ then x_0 is locally feasible and minimize g_G over the rate budget. Since x_0 only depends on the query-side nearest neighobur problem. Hence, it is independent of r.

Now wuppose $\lambda_+ \in (0,1)$, Lemma 11 we know $g_l(x_\lambda)$ is continuous and strictly decreasing in λ. We then have

$$g_l(x_{\lambda_+}) = s_*(\alpha; \rho_s)\,,$$

and $\lambda_+ < 1$. Theorem 8 proved the x_{λ_+} is the greedy optimizer. Therefore the next norm is

$$\alpha_+ = \alpha_{\lambda_+}(1 - m_{\alpha_{\lambda_+}})^2 + m_{\alpha_{\lambda_+}, NN}\,, \quad \alpha_{\lambda_+} = \lambda_+\alpha + (1-\lambda_+) - \lambda_+(1-\lambda_+)r\,.$$

Then we have

$$\frac{\partial \alpha_+}{\partial r} = \frac{\partial \alpha_+}{\partial \alpha_{\lambda_+}} \frac{\partial \alpha_{\lambda_+}}{\partial \lambda_+} \frac{\partial \lambda_+}{\partial r} = -\frac{(1-\lambda)(1-m_\lambda)(1-\theta)}{r + 2m_\lambda(\alpha - \theta^2)} < 0\,.$$

When $w \ge w_0$ with Lemma 27 we can see $\phi(\alpha, \theta) \in [\frac{\theta^2}{2}, \frac{1}{2})$. So $\theta = 1$ is impossible. □

To understand which next norms are typical, we also need to know how likely the source states are. We therefore keep the query fixed and consider the rate function of source.

Lemma 27. *The law of (α, θ) satisfies an LDP on $\mathbb{R}^2$ with speed d and good rate function*

$$\phi(\alpha, \theta) := \begin{cases} \frac{1}{2}[\alpha - 1 - \log(\alpha - \theta^2)], & \alpha > \theta^2\,, \\ \infty, & \alpha \le \theta^2\,. \end{cases}$$

Proof. Rotate the query to the first coordinate $q = \sqrt{d}e_1$ then write $v = \theta\sqrt{d}e_1 + v_\perp$. Then we have

$$\alpha = \theta^2 + \frac{\|v_\perp\|^2}{d}\,.$$

Use Theorem 4 and Lemma 6 gives the result. □

Definition 8. *For all feasible state and $\varphi > 0$, define*

$$I_{nav}(\varphi) := \inf\{\phi(\alpha, \theta) : \alpha_+(\alpha, r) = \varphi\}.$$

We then define the navigable hub region is

$$\mathcal{H}_{nav} := \left\{\varphi \in \mathcal{H}_{\text{geo}} : I_{nav}(\varphi) < \frac{1}{w}.\right\}$$

Only source states with rate below $1/w$ contribute to $\mathcal{H}_{nav}$. We use this restriction to bound the resulting output norms. The next lemma shows that the navigation region lies below the typical norm level and is bounded away from the smallest dataset norm level.

Lemma 28. *Let $w \geq w_0$. Then for every feasible state we have*

$$\inf \mathcal{H}_{nav} > \alpha_{min}\,.$$

Proof. For the lower bound we consider

$$\begin{aligned}\phi(\alpha, \theta) =& \alpha - 1 - \log(\alpha - \theta^2) \leq \frac{2}{w}\\ \alpha - \theta^2 \geq& \exp(\alpha - 1 - \frac{2}{w})\\ \alpha - \theta^2 \geq& \exp(\alpha_{min} - 1 - \frac{2}{w})\end{aligned}$$

Again, recall the definition of α_{min} is $\alpha_{min} - 1 - \log(\alpha_{min}) = \frac{2}{w}$. We have

$$\alpha - \theta^2 \geq \alpha_{min}\,.$$

Now recall

$$\alpha_{\lambda_+} = \frac{\|\lambda_+ v + (1-\lambda_+)q\|^2}{d} = \begin{pmatrix}\lambda & 1-\lambda\end{pmatrix}\begin{pmatrix}\alpha & \theta\\ \theta & 1\end{pmatrix}\begin{pmatrix}\lambda\\ 1-\lambda\end{pmatrix}.$$

Thus we have two inequality of the determinant of the middle matrix and the trace

$$\alpha - \theta^2 \geq \alpha_{min}, \quad \alpha + 1 \leq \alpha_{max} + 1\,.$$

Then for for every $\lambda_+ \in [0, 1]$ we have

$$\alpha_{\lambda_+} \geq \frac{\alpha_{min}}{\alpha_{max} + 1}[\lambda_+^2 + (1-\lambda_+)^2] > 0$$

Now recall $\frac{d}{d\gamma} m_{\gamma, NN} > 0$ we can obtain $\frac{d\alpha_{\lambda_+}}{d\alpha} > 0$. We also have $\alpha_+ = \alpha_{min}$ when $\alpha_{\lambda_+} = 0$. Therefore

$$\alpha_+ = \alpha_{\lambda_+}(1 - m_{\lambda_+})^2 + m_{\lambda_+} > \alpha_{min}\,.$$

Moreover, it is bound away from α_{min}. □

The typical starting points are powerful. Base on the Lemma 7 we know that the typical source is when $\alpha = 1$. Besides, when the dimension gets high, almost all vector will be perpendicular. This gives a intuitive view of navigable region. The most frequent visit norm stripe show deter by the typical norm 1 and typical angle $\cos(\theta) = 0$

We can now compare the geometric hub region with the one-step navigation region. The next theorem shows that the typical output norm lies in the geometric hub region. However, geometric hub norms sufficiently close to α_{min} lie outside the navigation region.

Theorem 11. *There exist a unique minimizer α_{nav} for I_{nav}*

$$\alpha_{nav} = \alpha_+(1, 2)\,, \quad I_{nav}(\alpha_{nav}) = 0\,, \quad I_{nav}(\varphi) > 0 \text{ for every } \varphi > 0,\, \varphi \neq \alpha_{nav}\,,$$

Moreover

$$\alpha_{min} < \alpha_{nav} < \alpha_h\,, \quad \alpha_{nav} \in \mathcal{H}_{nav}\,,$$

and

$$\inf \mathcal{H}_{nav} > \alpha_{min}\,, \sup \mathcal{H}_{nav} < \alpha_h\,.$$

Moreover, there exist a non-empty interval

$$(\alpha_{min}, \inf \mathcal{H}_{nav}) \subseteq \mathcal{H}_{geo} \backslash \mathcal{H}_{nav}\,.$$

Proof. First we can see the map $(\alpha, \theta) \mapsto \alpha_+(\alpha, r)$ is continuous on $\alpha > \theta^2$ then with Lemma 26,Lemma 27 and Theorem 4 we claim $\alpha_+(\alpha, r)$ satisfies and LDP with speed d and good rate function I_{nav}. Then the idea of Definition 8 is clear, fix a query q then for all possible source state where the greedy search and move to.

To obtain the minimizer following rewrite is useful (recall Theorem 6!)

$$\phi(\alpha, \theta) = \frac{\theta^2}{2} + I_0(\alpha - \theta^2)\,.$$

This shows that with $\phi \geq 0$ the unique zero is $(\alpha, \theta) = (1, 0)$ which is equivalent to $(\alpha, r) = (1, 2)$. Define

$$\alpha_{nav} := \alpha_+(1, 2)\,.$$

Then directly from the definition of I_{nav} we have

$$I_{nav}(\alpha_{nav}) \leq \phi(1, 0) = 0\,.$$

Since I_{nav} is a rate function which means it is non-negative, hence

$$I_{nav}(\alpha_{nav}) = 0\,.$$

As the unique zero of ϕ is $(1, 0)$, we can say α_{nav} is the unique minimizer and

$$I_{nav}(\varphi) > 0, \quad \text{for every } \varphi \neq \alpha_{nav}\,.$$

Then Lemma 28 gives

$$\inf \mathcal{H}_{nav} > \alpha_{min}.$$

What remain is to compare α_{nav} with the geometric hub boundary α_h. Recall Lemma 23 we have

$$\alpha_{flip} = (1 - m_{flip})^2 + m_{flip}\,, \quad m_{flip} := m_1(s_*(1; \rho_s))\,.$$

By considering two case in Lemma 26 we will show

$$\alpha_{nav} < \alpha_{flip}\,.$$

If $\lambda_+ = 0$, then we have $\alpha_{\lambda_+} = 0$. This gives $x_+ = x_0$ and $\alpha_+ = 1$. As x_0 is locally feasible,

$$g_l(x_0) = 1 + \alpha_{nav} \leq s_*(1; \rho_s) = m_{flip}^2 + m_{flip}.$$

With $m_{flip} \in (0, 1)$ we have

$$\alpha_{nav} \leq m_{flip}^2 + m_{flip} - 1 < m_{flip}^2 - m_{flip} + 1 = \alpha_{flip}\,.$$

Now we consider the $\lambda_+ \in (0, 1)$. Then the local constraint is on the boundary. We then have

$$g_l(x_+) = s_*(1; \rho_s) = m_{flip}^2 - m_{flip}\,.$$

At $(\alpha, \theta) = (1, 0)$, by Theorem 8 we have

$$a_+ = \lambda_+(1 - m_+)\,, \quad b_+ = (1 - \lambda_+)(1 - m_+)\,.$$

Then

$$\begin{aligned} m_{flip}^2 - m_{flip} &= (a_+ - 1)^2 + b_+^2 + m_+ \\ &= (m_+ + b_+)^2 + (m_+ + b_+) - b_+(1 - b_+) \\ &\leq (m_+ + b_+)^2 + (m_+ + b_+) \qquad (b_+ \in (0, 1))\,. \end{aligned}$$

This gives

$$m_{flip} < m_+ + b_+ = 1 - a_+\,.$$

We can rewrite the $g_l(x_+)$ in α_{nav} way

$$g_l(x_+) = \alpha_{nav} + 1 - 2a_+\,.$$

Now we have

$$\begin{aligned}\alpha_{nav} &= m_{flip}^2 - m_{flip} - 1 + 2a_+ \\ &< m_{flip}^2 - m_{flip} + 1 = \alpha_{flip}\,.\end{aligned}$$

Combine both case we have

$$\alpha_{nav} < \alpha_{flip}\,.$$

Recall Lemma 24 we have

$$I_{in}(\alpha_h) = \frac{1-\rho_s}{w}\,, \quad I_0(\alpha_{flip}) + I_{in}(\alpha_{flip}) = \frac{1-\rho_s}{w}\,.$$

Since $\alpha_{flip} < 1$ we have

$$I_{in}(\alpha_{flip}) < \frac{1-\rho_s}{w} \Rightarrow \alpha_{flip} < \alpha_h\,.$$

Combinning all the results gives

$$\alpha_{min} < \alpha_{nav} < \alpha_h \implies \alpha_{nav} \in \mathcal{H}_{nav}\,.$$

Finally, by Lemma 28 we have

$$\inf \mathcal{H}_{nav} > \alpha_{min}\,,$$

while Lemma 24 givs

$$\alpha_h > \alpha_{min}\,.$$

Therefore

$$(\alpha_{min}, \inf \mathcal{H}_{nav})$$

is a non-empty interval which finish the proof. □

E Spherical Uniform Distribution

In this section we apply (almost) same technique in Gaussian distribution to Spherical uniform distribution. We only consider the continuous one-step analysis. Through out this section let the spherical data be uniform on $\sqrt{d}\,S^{d-1}$ and take the query with $\|q\|^2/d = 1$. Since every source also has norm one, rotate the frame so that

$$v = \sqrt{d}\,e_1, \qquad q = \sqrt{d}\left(ze_1 + \sqrt{1-z^2}\,e_2\right),$$

where

$$z = \frac{\langle v, q\rangle}{d}, \qquad r = \frac{\|v-q\|^2}{d} = 2(1-z).$$

We use the same first two coordinates as in Part 1:

$$A_d = \frac{\langle U, e_1\rangle}{\sqrt{d}}, \qquad B_d = \frac{\langle U, e_2\rangle}{\sqrt{d}}.$$

The only difference is that now U lies on the sphere, so its remaining squared energy is exactly $1 - A_d^2 - B_d^2$. Hence the local term and greedy term become

$$g_l(a,b) = 2(1-a), \quad g_G(a,b) = 2\left(1 - za - \sqrt{1-z^2}\,b\right).$$

E.1 ρ_{uni} for Spherical Uniform distribution

We first determine the probability cost of a candidate's coordinates. The same rate function then gives the lower-tail distance rate used to obtain the nearest-neighbor scale and the local threshold.

Lemma 29. *The law of (A_d, B_d) satisfies an LDP on $\mathbb{R}^2$ with speed d and good rate function*

$$J_{uni}(a,b) = \begin{cases} -\frac{1}{2}\log(1-a^2-b^2), & a^2+b^2<1, \\ +\infty, & a^2+b^2 \geq 1\,. \end{cases}$$

We obtain the distance rate as

$$I_{uni}(s) = \begin{cases} -\frac{1}{2}\log(1-(1-\frac{s}{2})^2), & s\in(0,2), \\ +\infty, & otherwise. \end{cases}$$

Therefore we have

$$s_{uni,NN} = 2\left(1-\sqrt{1-\exp\left(-\frac{2}{w}\right)}\right)$$

as the nearest neighbour scale and the threshold is

$$s_{uni,*}(\rho_s) = 2\left(1-\sqrt{1-\exp\left(-2\frac{1-\rho_s}{w}\right)}\right).$$

Moreover,

$$J_{uni}(a,b) \leq \frac{1}{w} \iff a^2+b^2 \leq (1-\frac{s_{uni,NN}}{2})^2$$

Proof. For a uniform point on $\sqrt{d}\,S^{d-1}$, the first two normalized coordinates have density

$$f(a,b) = \frac{d-2}{\pi}(1-a^2-b^2)^{(d-4)/2}\,, \quad a^2+b^2<1\,.$$

Then all technique and relate definition to find such terms can be found in Theorem 5,Theorem 6,Definition 6. □

The spherical navigate family minimizes the same weighted combination of local and query distances as in the Gaussian model(we call it imaginary source). Its optimizer has the following explicit form.

Lemma 30. *Fix a spherical state with*

$$r \in (s_{uni,NN}, 4]\,, \quad z = 1-\frac{r}{2}\,.$$

For $\lambda\in[0,1]$, use the same navigate family in Theorem 8, we have

$$\alpha_{uni,\lambda} = 1-\lambda(1-\lambda)r\,.$$

If $0<r<4$ then the unique solution of

$$\min[\lambda g_l(a,b)+(1-\lambda)g_G(a,b)] \quad \textit{s.t. } J_{uni}(a,b) \leq \frac{1}{w}\,,$$

is

$$x_{uni,\lambda} = \frac{1-s_{uni,NN}/2}{\sqrt{\alpha_{uni,\lambda}}}(\lambda+(1-\lambda)z, (1-\lambda)\sqrt{1-z^2})\,.$$

Then in such spherical uniform navigate family, $g_l(x_{uni,\lambda})$ is strictly decreasing and $g_G(x_{uni,\lambda})$ increasing in λ.

Moreover let

$$s_{uni,crit}(r) := \min\left\{g_l(a,b) : J_{uni}(a,b) \leq \frac{1}{w},\ g_G(a,b) \leq r\right\}.$$

Then

$$s_{uni,crit}(r) = \begin{cases} s_{uni,NN}, & r \geq 2, \\ 2\left[1-z^2-\sqrt{(1-z^2)\left(\left(1-\frac{s_{uni,NN}}{2}\right)^2-z^2\right)}\right], & s_{uni,NN}<r<2, \end{cases}$$

where $z = 1-r/2$.

Proof. Check **Section B**. All technique are same, consider the end point, split into two case. The good things is in Spherical uniform dataset the local threshold is fixed so we can skip a lot of properties discussion. □

To obtain the required graph degree, we convert the critical local threshold into a degree exponent. Unlike the Gaussian case, there is no source-norm optimization: every source uses the same local threshold.

Theorem 12. *Let $c > 1$. In the non-trivial case $c^2 s_{uni,NN} < 2$, write $z := 1 - \frac{c^2 s_{uni,NN}}{2}$. Then the critical degree exponent for spherical uniform distribution is*

$$\rho_{uni}(c,w) = \begin{cases} 1 + \frac{w}{2}\log\left(1 - \left[z^2 + \sqrt{(1-z^2)\left(\left(1 - \frac{s_{uni,NN}}{2}\right)^2 - z^2\right)}\right]^2\right), & \text{if } c^2 s_{uni,NN} \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

Then for $\rho_u \in (0,1)$, strict progress exists from every spherical non-target state

$$r \ge c^2 s_{uni,NN}$$

if and only if

$$\rho_u > \rho_{uni}(c,w).$$

Proof. Again, check Theorem 9 for the technique. Unlike the Gaussian case, there is no source-norm interval to optimize over because every source has norm one and uses the same threshold. Thus the worst non-target state is simply the target boundary $r = c^2 s_{uni,NN}$, and substituting $s_{uni,crit}$ into the explicit spherical threshold calibration directly gives ρ_{uni}. □

We now compare the critical degree exponents as the approximation factor approaches one. The next theorem shows that their ordering at w_0 is reversed for sufficiently large w. By continuity, the two exponents must therefore intersect.

Theorem 13. *Let $w \ge w_0$, and let $s_0(w)$ be defined as in Lemma 16. Write*

$$\rho_{crit}(1_+, w) := \lim_{c \downarrow 1} \rho_{crit}(c,w) = 1 - w I_{\alpha_{min}}\big(s_0(w)\big),$$
$$\rho_{uni}(1_+, w) := \lim_{c \downarrow 1} \rho_{uni}(c,w) = \frac{w}{2}\log\big(2 - e^{-2/w}\big).$$

At $w = w_0$, we have

$$\rho_{crit}(1_+, w_0) > \rho_{uni}(1_+, w_0).$$

As $w \to \infty$,

$$\rho_{crit}(1_+, w) = \frac{8}{9} + O(w^{-1/2}), \qquad \rho_{uni}(1_+, w) = 1 - \frac{2}{w} + O(w^{-2}).$$

Consequently,

$$\rho_{crit}(1_+, w) < \rho_{uni}(1_+, w)$$

for all sufficiently large w. In particular, there exists at least one $w_\flat > w_0$ such that

$$\rho_{crit}(1_+, w_\flat) = \rho_{uni}(1_+, w_\flat).$$

Proof. The Gaussian endpoint formula follows from Lemma 16. For the spherical case, letting $c \downarrow 1$ in Theorem 12 gives

$$\begin{aligned} \rho_{uni}(1_+, w) &= 1 + \frac{w}{2}\log\left(1 - \left(1 - e^{-2/w}\right)^2\right) \\ &= \frac{w}{2}\log\big(2 - e^{-2/w}\big). \end{aligned}$$

We first compare the exponents at w_0. Until this comparison is complete, all quantities α_{min}, $m_{1,\mathrm{NN}}$, and $s_{\mathrm{NN}}(1)$ are evaluated at $w = w_0$. Let $m = m_{1,\mathrm{NN}}$ for convenience. Their defining equations give

$$\alpha_{min} + m^2 + m = 1, \qquad \frac{2}{w_0} = m^2 - m - \log m.$$

Since the Gaussian exponent is $1 - w_0 I_{\alpha_{min}}(s_0(w_0))$, it suffices to obtain a suitable upper bound on this rate. Consider the candidate

$$\left((1-m)\sqrt{\alpha_{min}},\ 0,\ m + (1-m)^2(1-\alpha_{min})\right).$$

Its third coordinate is positive. Moreover, using $s_{\mathrm{NN}}(1) = 1 - \alpha_{min}$, its local distance is

$$\begin{aligned}
&\alpha_{min} m^2 + m + (1-m)^2(1-\alpha_{min}) \\
&\quad = m^2 + (1-m)s_{\mathrm{NN}}(1) + \alpha_{min} m \\
&\quad = s_0(w_0).
\end{aligned}$$

Evaluating the variational representation of $I_{\alpha_{min}}$ at this candidate therefore gives

$$\begin{aligned}
I_{\alpha_{min}}\big(s_0(w_0)\big) &\le \frac{1}{2}\left[m^2 - m - \log\left(m + (1-m)^2(1-\alpha_{min})\right)\right] \\
&= \frac{1}{w_0} - \frac{1}{2}\log\left(1 + (1-m)^2(1+m)\right),
\end{aligned}$$

where the last equality uses $1 - \alpha_{min} = m + m^2$.

Since $0 < m < 1$, we have

$$e^{m(1-m)} > 1 + m(1-m).$$

Hence

$$\begin{aligned}
1 + (1-m)^2(1+m) &= 2 - m\left[1 + m(1-m)\right] \\
&> 2 - m e^{m(1-m)} \\
&= 2 - e^{-2/w_0}.
\end{aligned}$$

Both sides are positive. Combining this inequality with the preceding rate bound yields

$$\begin{aligned}
\rho_{crit}(1_+, w_0) &\ge \frac{w_0}{2}\log\left(1 + (1-m)^2(1+m)\right) \\
&> \frac{w_0}{2}\log\big(2 - e^{-2/w_0}\big) \\
&= \rho_{uni}(1_+, w_0).
\end{aligned}$$

We next consider $w \to \infty$. Expanding the spherical endpoint formula gives

$$\log\big(2 - e^{-2/w}\big) = \frac{2}{w} - \frac{4}{w^2} + O(w^{-3}),$$

and therefore

$$\rho_{uni}(1_+, w) = 1 - \frac{2}{w} + O(w^{-2}).$$

Together with Lemma 17, this shows that

$$\lim_{w\to\infty}\left[\rho_{uni}(1_+, w) - \rho_{crit}(1_+, w)\right] = \frac{1}{9} > 0.$$

Thus the Gaussian exponent is smaller for all sufficiently large w.

Finally, the defining equations for α_{min} and m have continuous, strictly monotone left-hand sides on $(0,1)$, so their solutions depend continuously on w. The endpoint formulas then show that both exponents are continuous on $[w_0, \infty)$. Their difference has opposite signs at w_0 and at all sufficiently large w. The intermediate value theorem therefore gives an intersection $w_\flat > w_0$. □