Supplementary Materials

A Analysis of spherical caps

In this section, we formulate some technical results on the volumes of spherical caps and their intersections, which we extensively use in the proofs. Although they are similar to those formulated in [1], it is crucial for our problem that parameters defining spherical caps depend on d and may tend to zero (both in dense and sparse regimes), while the results in [1] hold only for fixed parameters. Also, in Lemma 2, we extend the corresponding result from [1], as discussed further in this section.

A.1 Volumes of spherical caps

Let us denote by μ the Lebesgue measure over $\mathbb{R}^{d+1}$. By $C_{\mathbf{x}}(\gamma)$ we denote a spherical cap of height γ centered at $\mathbf{x} \in \mathcal{S}^d$, i.e., $\{\mathbf{y} \in \mathcal{S}^d : \langle \mathbf{x}, \mathbf{y} \rangle \geq \gamma\}$; $C(\gamma) = \mu\left(C_{\mathbf{x}}(\gamma)\right)$ denotes the volume of a spherical cap of height γ. Recall that throughout the paper for any variable γ, $0 \leq \gamma \leq 1$, we let $\hat{\gamma} := \sqrt{1-\gamma^2}$. We prove the following lemma.

Lemma 1. *Let $\gamma = \gamma(d)$ be such that $0 \leq \gamma \leq 1$. Then*

$$\Theta\left(d^{-1/2}\right) \hat{\gamma}^d \leq C\left(\gamma\right) \leq \Theta\left(d^{-1/2}\right) \hat{\gamma}^d \min\left\{d^{1/2}, \frac{1}{\gamma}\right\} .$$

Proof. In order to have similar reasoning with the proof of Lemma 2, we consider any two-dimensional plane containing the vector $\mathbf{x}$ defining the cap $C_{\mathbf{x}}(\gamma)$ and let p denote the orthogonal projection from $\mathcal{S}^d$ to this two-dimensional plane.

The first steps of the proof are similar to those in [1] (but note that we analyze $\mathcal{S}^d$ instead of S^{d-1}, which leads to slightly simpler expressions). Consider any measurable subset U of the two-dimensional unit ball, then the volume of the preimage $p^{-1}(U)$ (relative to the volume of $\mathcal{S}^d$) is:

$$I(U) = \int_{r,\phi \in U} \frac{\mu(S^{d-2})}{\mu(S^d)} \left(\sqrt{1-r^2}\right)^{d-3} r\, dr\, d\phi\,.$$

We define a function $g(r) = \int_{\phi:(r,\phi)\in U} d\phi$, then we can rewrite the integral as

$$I(U) = \frac{(d-1)}{4\,\pi} \int_0^1 \left(1-r^2\right)^{(d-3)/2} g(r)\, dr^2\,.$$

Let $U = p\left(C_{\mathbf{x}}(\gamma)\right)$, then, using $t = \left(1-r^2\right)/\hat{\gamma}^2$, where $\hat{\gamma} = \sqrt{1-\gamma^2}$, we get

$$C(\gamma) = \frac{(d-1)}{4\,\pi} \int_\gamma^1 \left(1-r^2\right)^{(d-3)/2} g(r)\, dr^2 = \frac{(d-1)\,\hat{\gamma}^{d-1}}{4\,\pi} \int_0^1 g\left(\sqrt{1-\hat{\gamma}^2 t}\right) t^{(d-3)/2}\, dt\,. \tag{1}$$

Note that from Equation (1) we get thvolume of a hemisphere is $C(0) = 1/2$, since $g(r) = \pi$ for all r in this case and $\hat{\gamma} = 1$.

Now we consider an arbitrary $\gamma \geq 0$ and note that $g(r) = 2\arccos(\gamma/r)$ (see Figure 1). So, we obtain

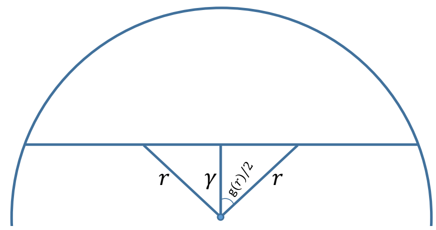


Figure 1: $g(r)$

$$C(\gamma) = \frac{(d-1)\,\hat{\gamma}^{d-1}}{2\,\pi} \int_0^1 \arccos\left(\frac{\gamma}{\sqrt{1-\hat{\gamma}^2 t}}\right) t^{\frac{d-3}{2}}\,dt$$
$$= \frac{(d-1)\,\hat{\gamma}^{d-1}}{2\,\pi} \int_0^1 \arcsin\left(\hat{\gamma}\sqrt{\frac{1-t}{1-\hat{\gamma}^2 t}}\right) t^{(d-3)/2}\,dt\,.$$

Now we note that $x \le \arcsin(x) \le x \cdot \pi/2$ for $0 \le x \le 1$, so

$$C(\gamma) = \Theta\left(d\right) \hat{\gamma}^d \int_0^1 \sqrt{\frac{1-t}{1-\hat{\gamma}^2 t}} \cdot t^{(d-3)/2}\,dt\,.$$

Finally, we estimate

$$\sqrt{1-t} \le \sqrt{\frac{1-t}{1-\hat{\gamma}^2 t}} \le \min\left\{1, \sqrt{\frac{1-t}{1-\hat{\gamma}^2}}\right\}. \tag{2}$$

So, the lower bound is

$$C(\gamma) \ge \Theta\left(d\right) \hat{\gamma}^d\, \mathrm{B}\left(\frac{3}{2}, \frac{d-1}{2}\right) = \Theta\left(d\right) \hat{\gamma}^d \left(\frac{d-1}{2}\right)^{-3/2} = \Theta\left(d^{-1/2}\right) \hat{\gamma}^d\,.$$

The upper bounds are

$$C(\gamma) \le \Theta\left(d\right) \hat{\gamma}^d \int_0^1 t^{(d-3)/2}\,dt = \Theta\left(1\right) \hat{\gamma}^d\,,$$
$$C(\gamma) \le \Theta\left(d\right) \hat{\gamma}^d \int_0^1 \sqrt{\frac{1-t}{1-\hat{\gamma}^2}} \cdot t^{(d-3)/2}\,dt = \Theta\left(d^{-1/2}\right) \frac{\hat{\gamma}^d}{\gamma}\,.$$

This completes the proof. □

A.2 Volumes of intersections of spherical caps

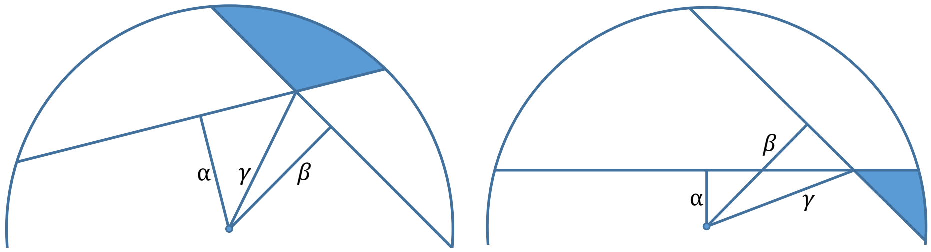
By $W_{\mathbf{x},\mathbf{y}}(\alpha, \beta)$ we denote the intersection of two spherical caps centered at $\mathbf{x} \in \mathcal{S}^d$ and $\mathbf{y} \in \mathcal{S}^d$ with heights α and β, respectively, i.e., $W_{\mathbf{x},\mathbf{y}}(\alpha, \beta) = \{\mathbf{z} \in \mathcal{S}^d : \langle \mathbf{z}, \mathbf{x}\rangle \ge \alpha, \langle \mathbf{z}, \mathbf{y}\rangle \ge \beta\}$. As for spherical caps, by $W(\alpha, \beta, \theta)$ we denote the volume of such intersection given that the angle between $\mathbf{x}$ and $\mathbf{y}$ is θ.

We analyze the volume of the intersection of two spherical caps $C_{\mathbf{x}}(\alpha)$ and $C_{\mathbf{y}}(\beta)$. In the lemma below we assume $\gamma \le 1$. However, it is clear that if $\gamma > 1$, then either the caps do not intersect (if $\alpha > \beta\cos\theta$ and $\beta > \alpha\cos\theta$) or the larger cap contains the smaller one.

Lemma 2. *Let $\gamma = \frac{\sqrt{\alpha^2+\beta^2-2\alpha\beta\cos\theta}}{\sin\theta}$ and assume that $\gamma \le 1$, then:*

(1) If $\alpha \le \beta\cos\theta$, then $C(\beta)/2 < W(\alpha, \beta, \theta) \le C(\beta)$ and

$$C(\beta) - W(\alpha, \beta, \theta) \le C_{u,\beta}\,\Theta\left(d^{-1}\right) \hat{\gamma}^d \min\left\{d^{1/2}, \frac{1}{\gamma}\right\};$$



(a) $\beta > \alpha\cos\theta$, $\alpha > \beta\cos\theta$ (b) $\alpha < \beta\cos\theta$

Figure 2: Intersection of spherical caps

(2) If $\beta \le \alpha\cos\theta$, then $C(\alpha)/2 < W(\alpha,\beta,\theta) \le C(\alpha)$ and

$$C(\alpha) - W(\alpha,\beta,\theta) \le C_{u,\alpha}\,\Theta\left(d^{-1}\right)\hat{\gamma}^d \min\left\{d^{1/2}, \frac{1}{\gamma}\right\};$$

(3) Otherwise,

$$(C_{l,\alpha} + C_{l,\beta})\,\Theta\left(d^{-1}\right)\hat{\gamma}^d \le W(\alpha,\beta,\theta) \le (C_{u,\alpha} + C_{u,\beta})\,\Theta\left(d^{-1}\right)\hat{\gamma}^d \min\left\{d^{1/2}, \frac{1}{\gamma}\right\};$$

where

$$C_{l,\alpha} = \frac{\alpha\left(\hat{\alpha}\,\sin\theta - |\beta - \alpha\,\cos\theta|\right)}{\gamma\,\hat{\gamma}\,\sin\theta}\,, \quad C_{l,\beta} = \frac{\beta\left(\hat{\beta}\,\sin\theta - |\alpha - \beta\,\cos\theta|\right)}{\gamma\,\hat{\gamma}\,\sin\theta}\,,$$
$$C_{u,\alpha} = \frac{\hat{\gamma}\,\alpha\,\sin\theta}{\gamma|\beta - \alpha\,\cos\theta|}\,, \quad C_{u,\beta} = \frac{\hat{\gamma}\,\beta\,\sin\theta}{\gamma|\alpha - \beta\,\cos\theta|}\,.$$

The cases considered in this lemma are illustrated in Figure 2.

This lemma differs from Lemma 2.2 in [1] by, first, allowing the parameters α, β, θ depend on d and, second, considering the cases (1) and (2), which are essential for the proofs. Namely, we use the lower bound in (3) to show that with high probability we can make a step of the algorithm since the intersection of some spherical caps is large enough (Figure 2a); we use the upper bounds in (1) and (2) to show that at the final step of the algorithm we can find the nearest neighbor with high probability, since the volume of the intersection of some spherical caps is very close to the volume of one of them (Figure 2b), see the details further in the proof.

Proof. Consider the plane formed by the the vectors $\mathbf{x}$ and $\mathbf{y}$ defining the caps and let p denote the orthogonal projection to this plane. Let $U = p(W_{\mathbf{x},\mathbf{y}}(\alpha,\beta))$.

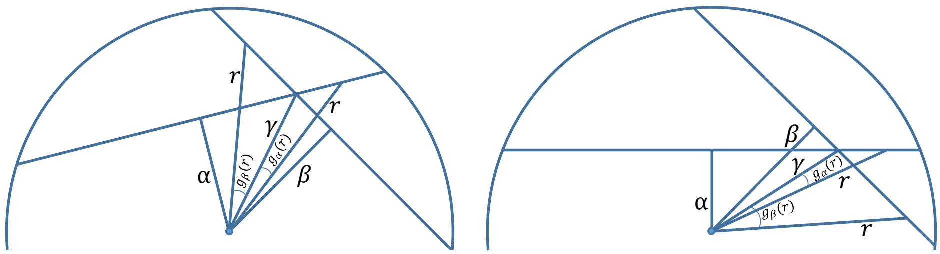
Denote by γ the distance between the origin and the intersection of chords bounding the projections of spherical caps. One can show that

$$\gamma = \sqrt{\frac{\alpha^2 + \beta^2 - 2\alpha\beta\cos\theta}{\sin^2\theta}}\,.$$

If $\alpha \le \beta\cos\theta$, it is easy to see that $W(\alpha,\beta,\theta) > \frac{1}{2}C(\beta)$, since more than a half of $C_{\mathbf{y}}(\beta)$ is covered by the intersection (see Figure 2b). Similarly, if $\beta \le \alpha\cos\theta$, then $W(\alpha,\beta,\theta) > \frac{1}{2}C(\alpha)$. Now we move to the proof of (3) and will return to (1) and (2) after that.

If $\cos\theta < \frac{\alpha}{\beta}$ and $\cos\theta < \frac{\beta}{\alpha}$, then we are in the situation shown on Figure 2a and the distance between the intersection of spherical caps and the origin is γ. As in the proof of Lemma 1, denote $g(r) = \int_{\phi:(r,\phi)\in U} d\phi$, then the relative volume of $p^{-1}(U)$ is (see Equation (1)):

$$W(\alpha,\beta,\theta) = \frac{(d-1)\,\hat{\gamma}^{d-1}}{4\,\pi}\int_0^1 g\left(\sqrt{1-\hat{\gamma}^2 t}\right)t^{\frac{d-3}{2}}\,dt\,.$$



(a) $\beta > \alpha\cos\theta$, $\alpha > \beta\cos\theta$ (b) $\alpha < \beta\cos\theta$

Figure 3: $g_\alpha(r)$ and $g_\beta(r)$

The function $g(r)$ can be written as $g_\alpha(r) + g_\beta(r)$, where (see Figure 3a)

$$g_\alpha(r) = \arccos\left(\frac{\alpha}{r}\right) - \arccos\left(\frac{\alpha}{\gamma}\right),$$
$$g_\beta(r) = \arccos\left(\frac{\beta}{r}\right) - \arccos\left(\frac{\beta}{\gamma}\right).$$

Accordingly, we can write $W(\alpha,\beta,\theta) = W_\alpha(\alpha,\beta,\theta) + W_\beta(\alpha,\beta,\theta)$.

Let us estimate $g_\alpha\left(\sqrt{1-\hat\gamma^2 t}\right)$:

$$\begin{aligned}
g_\alpha\left(\sqrt{1-\hat\gamma^2 t}\right) &= \arcsin\left(\sqrt{1-\frac{\alpha^2}{1-\hat\gamma^2 t}}\right) - \arcsin\left(\sqrt{1-\frac{\alpha^2}{\gamma^2}}\right) \\
&= \arcsin\left(\sqrt{\left(1-\frac{\alpha^2}{1-\hat\gamma^2 t}\right)\frac{\alpha^2}{\gamma^2}}\sqrt{\left(1-\frac{\alpha^2}{\gamma^2}\right)\frac{\alpha^2}{1-\hat\gamma^2 t}}\right) \\
&= \Theta\left(\frac{\alpha\sqrt{\gamma^2-\alpha^2}}{\gamma\sqrt{1-\hat\gamma^2 t}}\right)\left(\sqrt{1+\frac{\hat\gamma^2(1-t)}{\gamma^2-\alpha^2}}-1\right).
\end{aligned}$$

Note that

$$\left(\sqrt{1+\frac{\hat\gamma^2}{\gamma^2-\alpha^2}}-1\right)(1-t) \le \sqrt{1+\frac{\hat\gamma^2(1-t)}{\gamma^2-\alpha^2}}-1 \le \frac{\hat\gamma^2}{2\left(\gamma^2-\alpha^2\right)}(1-t).$$

Now, we can write the lower bound for $W_\alpha(\alpha,\beta,\theta)$. Let

$$C_{l,\alpha} = \left(\sqrt{1+\frac{\hat\gamma^2}{\gamma^2-\alpha^2}}-1\right)\frac{\alpha\sqrt{\gamma^2-\alpha^2}}{\gamma\,\hat\gamma} = \frac{\alpha\left(\hat\alpha\,\sin\theta - |\beta-\alpha\,\cos\theta|\right)}{\gamma\,\hat\gamma\,\sin\theta},$$

$C_{l,\beta}$ can be obtained by swapping α and β.

Then the lower bound is

$$\begin{aligned}
W(\alpha,\beta,\theta) &\ge \Theta(d)\,\hat\gamma^d\left(C_{l,\alpha}+C_{l,\beta}\right)\int_0^1 \frac{1-t}{\sqrt{1-\hat\gamma^2 t}}t^{(d-3)/2}\,dt \\
&\ge \Theta(d)\,\hat\gamma^d\left(C_{l,\alpha}+C_{l,\beta}\right)\int_0^1 (1-t)\,t^{(d-3)/2}\,dt = \Theta(d^{-1})\,\hat\gamma^d\left(C_{l,\alpha}+C_{l,\beta}\right).
\end{aligned}$$

Now we define $C_{u,\alpha}$ (and, similarly, $C_{u,\beta}$) as

$$C_{u,\alpha} = \frac{\hat\gamma}{\left(\gamma^2-\alpha^2\right)}\cdot\frac{\alpha\sqrt{\gamma^2-\alpha^2}}{\gamma} = \frac{\hat\gamma\,\alpha\,\sin\theta}{\gamma\,|\beta-\alpha\,\cos\theta|}.$$

Then

$$W(\alpha,\beta,\theta) \le \Theta(d)\,\hat{\gamma}^d\,(C_{u,\alpha}+C_{u,\beta})\,\cdot \int_0^1 \frac{1-t}{\sqrt{1-\hat{\gamma}^2 t}} t^{(d-3)/2}\,dt\,.$$

We use the upper bound

$$\frac{1-t}{\sqrt{1-\hat{\gamma}^2 t}} \le \min\left\{\sqrt{1-t}, \frac{1-t}{\gamma}\right\}$$

and obtain

$$W(\alpha,\beta,\theta) \le \Theta(d^{-1})\,\hat{\gamma}^d\,(C_{u,\alpha}+C_{u,\beta}) \min\left\{d^{\frac{1}{2}}, \frac{1}{\gamma}\right\},$$

which completes the proof of (3).

Now, let us finish the proof for (1) and (2). If $\alpha \le \beta\cos\theta$, then we are in a situation shown on Figure 2b. In this case, the bounds on $W(\alpha,\beta,\theta)$ are obvious. To estimate $C(\beta) - W(\alpha,\beta,\theta)$, we can directly follow the above proof for (3) and the only difference would be that $g(r) = g_\beta(r) - g_\alpha(r)$ instead of $g(r) = g_\alpha(r) + g_\beta(r)$. Note that we need only the upper bound and we simply say that $g(r) \le g_\beta(r)$. The proof for (2) is similar with $g(r) = g_\alpha(r) - g_\beta(r)$. $\square$

B Greedy search on plain NN graphs

B.1 Proof overview

Let α_M denote the height of a spherical cap defining $G(M)$. By $f = f(n) = (n-1)C(\alpha_M)$ we denote the expected number of neighbors of a given node in $G(M)$. Then, it is clear that the complexity of one step of graph-based search is $\Theta\left(f \cdot d\right)$ (with high probability), so for making k steps we need $\Theta\left(k \cdot f \cdot d\right)$ computations (see Section B.2.1). The number of edges in the graph is $\Theta\left(f \cdot n\right)$, so the space complexity is $\Theta\left(f \cdot n \cdot \log n\right)$ (see Section B.2.2).

To prove that the algorithm succeeds, we have to show that it does not get stuck in a local optimum until we are sufficiently close to q. If we take some point $\mathbf{x}$ with $\langle \mathbf{x}, q\rangle = \alpha_s$, then the probability of making a step towards q is determined by $W(\alpha_M, \alpha_s, \arccos\alpha_s)$. In all further proofs we obtain lower bounds for this value of the form $\frac{1}{n}g(n)$ with $1 \ll g(n) \ll n$. From this, we easily get that the probability of making a step is at least $1 - \left(1 - g(n)/n\right)^{n-1} = 1 - e^{-g(n)(1+o(1))}$.

A fact that will be useful in the proofs is that the value $W(\alpha_M, \alpha_s, \arccos\alpha_s)$ is a monotone function of α_s (see Section B.2.3). I.e., if we have a lower bound for some α_s, then for all smaller values we have this bound automatically.

By estimating the value $W(\alpha_M, \alpha_s, \arccos\alpha_s)$, we obtain (in further sections) that with probability $1 - o(1)$ we reach some point at distance at most $\arccos\alpha_s$ from q. Then, to achieve success, we may either jump directly to $\bar{\mathbf{x}}$ at the next step or to already have $\arccos\alpha_s \le cR$ if we are solving c, R-ANN.

To limit the number of steps, we additionally show that with a sufficiently large probability at each step we become "ε closer" to q. In the dense regime, it means that the sine of the angle between the current position and q becomes smaller by at least some fixed value.

Let us emphasize that several consecutive steps of the algorithm cannot be analyzed independently. Indeed, if at some step we moved from $\mathbf{x}$ to $\mathbf{y}$, then there were no points in $C_{\mathbf{x}}(\alpha_M)$ closer to q than $\mathbf{y}$ by the definition of the algorithm. Consequently, the intersection of $C_{\mathbf{x}}(\alpha_M)$, $C_{\mathbf{y}}(\alpha_M)$ and $C_q\left(\langle q, \mathbf{y}\rangle\right)$ contains no elements of the dataset. The closer $\mathbf{y}$ to $\mathbf{x}$ the larger this intersection. However, the fact that at each step we become at least "ε closer" to q allows us to bound the volume of this intersection and to prove that it can be neglected.

This is worth noting that in the proofs below we assume that the elements are distributed according to the Poisson point process on $\mathcal{S}^d$ with n being the *expected* number of elements. This makes the proofs more concise without changing the results since the distributions are asymptotically equivalent. Indeed, conditioning on the number of nodes in the Poisson process, we get the uniform distribution, and the number of nodes in the Poisson process is $\Theta(n)$ with high probability.

B.2 Auxiliary results

B.2.1 Time complexity

Let v be an arbitrary node of G and let $N(v)$ denote the number of its neighbors in G. Recall that $f = (n-1)C(\alpha_M)$.

Lemma 3. *With probability at lest $1 - \frac{4}{f}$ we have $\frac{1}{2}f \le N(v) \le \frac{3}{2}f$.*

Proof. The number of neighbors $N(v)$ of a node v follows Binomial distribution $\mathrm{Bin}(n-1, C(\alpha_M))$, so $\mathbb{E}N(v) = f$. From Chebyshev's inequality we get

$$\mathrm{P}\left(|N(v) - f| > \frac{f}{2}\right) \le \frac{4\,\mathrm{Var}(N(v))}{f^2} \le \frac{4}{f},$$

which completes the proof. □

To obtain the final time complexity of graph-based NN search, we have to sum up the complexities of all steps of the algorithm. We obtain the following result.

Lemma 4. *If we made k steps of the graph-based NNS, then with probability $1 - O\left(\frac{1}{k\,f}\right)$ the obtained time complexity is $\Theta\left(kfd\right)$.*

Proof. Although the nodes encountered in one iteration are not independent, the fact that we do not need to measure the distance from any point to q more than once allows us to upper bound the complexity by the random variable distributed according to $\mathrm{Bin}(k(n-1), C(\alpha_M))$. Then, we can follow the proof of Lemma 3 and note that one distance computation takes $\Theta(d)$.

To see that the lower bound is also $\Theta\left(kfd\right)$, we note that more than a constant number of steps are needed only for the dense regime. For this regime, we may follow the reasoning of Lemma 10 to show that the volume of the intersection of two consecutive balls is negligible compared to the volume of each of them. □

B.2.2 Space complexity

Lemma 5. *With probability $1 - O\left(\frac{1}{f\,n}\right)$ we have $\frac{1}{4}\,f\,n \le E(G) \le \frac{3}{4}\,f\,n$.*

Proof. The proof is straightforward.[1] For each pair of nodes, the probability that there is an edge between them equals $C\left(\alpha_M\right)$. Therefore, the expected number of edges is

$$\mathbb{E}(E(G)) = \binom{n}{2} C\left(\alpha_M\right) \,.$$

It remains to prove that $E(G)$ is tightly concentrated near its expectation. For this, we apply Chebyshev's inequality, so we have to estimate the variance $\mathrm{Var}(E(G))$. One can easily see that if we are given two pairs of nodes e_1 and e_2, then, if they are not the same (while one coincident node is allowed), then $\mathrm{P}(e_1, e_2 \in E(G)) = C(\alpha_M)^2$. Therefore,

$$\begin{aligned}\mathrm{Var}(E(G)) &= \sum_{e_1, e_2 \in \binom{\mathcal{D}}{2}} \mathrm{P}(e_1, e_2 \in E(G)) \; - (\mathbb{E}E(G))^2 \\ &= \sum_{\substack{e_1, e_2 \in \binom{\mathcal{D}}{2} \\ e_1 \neq e_2}} \mathrm{P}(e_1, e_2 \in E(G)) + \mathbb{E}E(G) - (\mathbb{E}E(G))^2 = \binom{n}{2} C\left(\alpha_M\right)\left(1 - C\left(\alpha_M\right)\right) \,.\end{aligned}$$

Applying Chebyshev's inequality, we get

$$\mathrm{P}\left(|E(G) - \mathbb{E}(E(G))| > \frac{E(G)}{2}\right) \le \frac{4\,\mathrm{Var}(E(G))}{E(G)^2} = \frac{4\left(1 - C(\alpha)\right)}{E(G)} \,. \tag{3}$$

From this, the lemma follows. □

[1] Similar proof appeared in, e.g., [4].

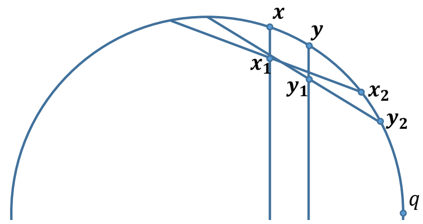


Figure 4: Monotonicity of $W(\alpha_M, \alpha_s, \arccos \alpha_s)$

It remains to note that if we store a graph as the adjacency lists, then the space complexity is $\Theta\left(E(G) \cdot \log n\right)$.

B.2.3 Monotonicity of $W(\alpha_M, \alpha_s, \arccos \alpha_s)$

Lemma 6. *$W(\alpha_M, \alpha_s, \arccos \alpha_s)$ is a non-increasing function of α_s.*

Proof. We refer to Figure 4, where two spherical caps of height α_M are centered at $\mathbf{x}$ and $\mathbf{y}$, respectively, and note that we have to compare "curved triangles" $\triangle\, \mathbf{x}\, \mathbf{x}_1\, \mathbf{x}_2$ and $\triangle\, \mathbf{y}\, \mathbf{y}_1\, \mathbf{y}_2$. Obviously, $\rho(\mathbf{x}, \mathbf{x}_2) = \rho(\mathbf{y}, \mathbf{y}_2)$, $\angle\, \mathbf{x}\, \mathbf{x}_2\, \mathbf{x}_1 = \angle\, \mathbf{y}\, \mathbf{y}_2\, \mathbf{y}_1$, but $\angle\, \mathbf{x}\, \mathbf{x}_1\, \mathbf{x}_2 < \angle\, \mathbf{y}\, \mathbf{y}_1\, \mathbf{y}_2$. From this and the spherical symmetry of $\mu(p^{-1}(\cdot))$ (p was defined in the proof of Lemma 2) the result follows. $\square$

B.3 Proof of Theorem 1 (greedy search in dense regime)

Recall that for dense datasets ($d = \log n/\omega$), it is convenient to operate with radii of spherical caps (if α is a height of a spherical cap, then we say that $\hat{\alpha}$ is its radius). Let $\hat{\alpha}_1$ be the radius of a cap centered at a given point and covering its nearest neighbor, then we have $C(\alpha_1) \sim \frac{1}{n}$, i.e., $\hat{\alpha}_1 \sim n^{-\frac{1}{d}} = 2^{-\omega}$. We further let $\delta := 2^{-\omega}$.

We construct a graph $G(M)$ using spherical caps with radius $\hat{\alpha}_M = M\,\delta$. Then, from Lemma 1, we get $f = \Theta\left(n\, d^{-1/2} M^d \delta^d\right) = \Theta\left(d^{-1/2} M^d\right)$. So, the number of edges in $G(M)$ is $\Theta\left(d^{-1/2} \cdot M^d \cdot n\right)$ and the space complexity is $\Theta\left(d^{-1/2} \cdot M^d \cdot n \cdot \log n\right)$ (see Section B.2.2).

Let us now analyze the distance $\arccos \alpha_s$ up to which we can make steps towards the query q (with sufficiently large probability). This is stated in Lemma 9, but before that let us prove some auxiliary results.

First, let us analyze the behavior of the main term $\hat{\gamma}^d$ in $W(x, y, \arccos z)$ when $\hat{x}, \hat{y}, \hat{z} = o(1)$, which is the case for the considered situations in the dense regime.

Lemma 7. *If $\hat{x}, \hat{y}, \hat{z} = o(1)$, then $\hat{\gamma}$ defined in Lemma 2 for $W(x, y, \arccos z)$ is*

$$\hat{\gamma} \sim \frac{\sqrt{2\left(\hat{x}^2\hat{y}^2 + \hat{y}^2\hat{z}^2 + \hat{x}^2\hat{z}^2\right) - \left(\hat{x}^4 + \hat{y}^4 + \hat{z}^4\right)}}{2\hat{z}}.$$

Proof. By the definition, $\gamma = \sqrt{\frac{x^2+y^2-2xyz}{1-z^2}}$. Then

$$\hat{\gamma}^2 = 1 - \gamma^2 = \frac{\hat{x}^2 + \hat{y}^2 + \hat{z}^2 - 2 + 2\sqrt{\left(1-\hat{x}^2\right)\left(1-\hat{y}^2\right)\left(1-\hat{z}^2\right)}}{\hat{z}^2}$$
$$\sim \frac{2\left(\hat{x}^2\hat{y}^2 + \hat{y}^2\hat{z}^2 + \hat{x}^2\hat{z}^2\right) - \left(\hat{x}^4 + \hat{y}^4 + \hat{z}^4\right)}{4\hat{z}^2}.$$

$\square$

Now, we analyze $W(\alpha_M, \alpha_s, \arccos \alpha_s)$ and we need only the lower bound. Recall that we use the notation $\delta = 2^{-\omega}$.

Lemma 8. *Assume that $\hat{\alpha}_s = s\,\delta$ and $\hat{\alpha}_M = M\,\delta$.*

- *If $M \geq \sqrt{2}s$, then $W(\alpha_M, \alpha_s, \arccos \alpha_s) \geq \frac{1}{n} \cdot s^{d+o(d)}$.*
- *If $M < \sqrt{2}s$, then $W(\alpha_M, \alpha_s, \arccos \alpha_s) \geq \frac{1}{n} \cdot \left(M^2 - \frac{M^4}{4s^2}\right)^{d/2+o(d)}$.*

Proof. First, assume that $M \geq \sqrt{2}s$. In this case we have $\alpha_M < \alpha_s^2$, so we are under the conditions (1)-(2) of Lemma 2 (see Figure 2b) and, using Lemma 1, we get $W(\alpha_M, \alpha_s, \arccos \alpha_s) > \frac{1}{2}C(\alpha_s) = \Theta\left(d^{-1/2}s^d\delta^d\right) = \frac{1}{n}s^{d+o(d)}$.

If $M < \sqrt{2}s$, then, asymptotically, we have $\alpha_M > \alpha_s^2$, so the case (3) of Lemma 2 can be applied. Let us use Lemma 7 to estimate $\hat{\gamma}$:

$$\hat{\gamma}^2 = \delta^2 \left(M^2 - \frac{M^4}{4s^2}\right)(1 + o(1)) \,.$$

And now from Lemma 2 we get

$$W(\alpha_M, \alpha_s, \arccos \alpha_s) \geq C_l\, \Theta\left(d^{-1}\right) \hat{\gamma}^d \,,$$

where C_l corresponds to the sum of $C_{l,\alpha}$ and $C_{l,\beta}$ in Lemma 2. So, it remains to estimate C_l:

$$\begin{aligned} C_l &= \frac{\alpha_M\left(\hat{\alpha}_M\,\hat{\alpha}_s + \alpha_M\alpha_s - \alpha_s\right) + \alpha_s\left(\hat{\alpha}_s^2 + \alpha_s^2 - \alpha_M\right)}{\gamma\,\hat{\gamma}\,\hat{\alpha}_s} \\ &= \Theta\left(\frac{Ms\delta^2 + \sqrt{(1-M^2\delta^2)(1-s^2\delta^2)} - \sqrt{1-s^2\delta^2} + s^2\delta^2 + 1 - s^2\delta^2 - \sqrt{1-M^2\delta^2}}{\delta^2}\right) \\ &= \Theta(1) \cdot \frac{M(s - M/2)\delta^2 + \frac{1}{2}M^2\delta^2}{\delta^2} = \Theta(1) \,. \end{aligned}$$

Therefore,

$$W(\alpha_M, \alpha_s, \arccos \alpha_s) \geq \frac{1}{n} \cdot \left(M^2 - \frac{M^4}{4s^2}\right)^{d/2+o(d)} \,.$$

□

From Lemma 8, we can find the conditions for M and s to guarantee (with sufficiently large probability) making steps until we are in the cap of radius $s\delta$ centered at q. The following lemma gives such conditions and also guarantees that at each step we can reach a cap of a radius at least $\varepsilon\delta$ smaller for some constant $\varepsilon > 0$.

Lemma 9. *Assume that $s > 1$. If $M > \sqrt{2}s$ or $M^2 - \frac{M^4}{4s^2} > 1$, then there exists such constant $\varepsilon > 0$ that $W(\alpha_M, \alpha_s, \arcsin(\hat{\alpha}_s + \varepsilon\delta)) \geq \frac{1}{n}S^{d(1+o(1))}$ for some constant $S > 1$.*

Proof. First, let us take $\varepsilon = 0$. Then, the result directly follows from Lemma 8. We note that a value M satisfying $M^2 - \frac{M^4}{4s^2} > 1$ exists only if $s > 1$.

Now, let us demonstrate that we can take some $\varepsilon > 0$. The two cases discussed in Lemma 8 now correspond to $M^2 \geq s^2 + (s+\varepsilon)^2$ and $M^2 < s^2 + (s+\varepsilon)^2$, respectively. If $M > \sqrt{2}s$, then we can choose a sufficiently small ε that $M^2 \geq s^2 + (s+\varepsilon)^2$. Then, the result follows from Lemma 8 and the fact that $s > 1$. Otherwise, we have $M^2 < s^2 + (s+\varepsilon)^2$ and instead of the condition $M^2 - \frac{M^4}{4s^2} > 1$ we get (using Lemma 7)

$$\frac{2\left(M^2s^2 + s^2(s+\varepsilon)^2 + M^2(s+\varepsilon)^2\right) - \left(M^4 + s^4 + (s+\varepsilon)^4\right)}{4(s+\varepsilon)^2} > 1.$$

As this holds for $\varepsilon = 0$, we can choose a small enough $\varepsilon > 0$ that the condition is still satisfied. □

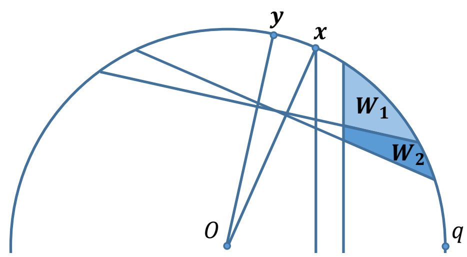


Figure 5: Dependence of consecutive steps

This lemma implies that if we are given M and s satisfying the above conditions, then we can make a step towards q since the expected number of nodes in the intersection of spherical caps is much larger than 1. Formally, we can estimate from below the values $g(n)$ for all steps of the algorithm by $g_{min}(n) = S^{d(1+o(1))}$. So, according to Section B.1, we can make each step with probability $1 - O\left(e^{-S^{d(1+o(1))}}\right)$. Moreover, each step reduces the radius of a spherical cap centered at q and containing the current position by at least $\varepsilon\delta$. As a result, the number of steps (until we reach some distance $\arccos \alpha_s$) is $O\left(\delta^{-1}\right) = O\left(2^{\omega}\right)$.

To estimate the overall success probability, we have to take into account that the consecutive steps of the algorithm are dependent. In Section B.1, it is explained how the previous steps of the algorithm may affect the current one: the fact that at some step we moved from $\mathbf{y}$ to $\mathbf{x}$ implies that there were no elements closer to q than $\mathbf{x}$ in a spherical cap centered at $\mathbf{y}$. However, we can show that this dependence can be neglected.

Lemma 10. *The dependence of the consecutive steps can be neglected and does not affect the analysis.*

Proof. The main idea is illustrated on Figure 5. Assume that we are currently at a point $\mathbf{x}$ with $\rho(\mathbf{x}, q) = \arcsin\left(\hat{\alpha}_s + \varepsilon\delta\right)$. Then, as in the proof of Lemma 9, we are interested in the volume $W(\alpha_M, \alpha_s, \arcsin\left(\hat{\alpha}_s + \varepsilon\delta\right))$, which corresponds to $W_1 + W_2$ on Figure 5. Assume that at the previous step we were at some point $\mathbf{y}$. Given that all steps are "longer than ε", the largest effect of the previous step is reached when $\mathbf{y}$ is as close to $\mathbf{x}$ as possible and $\mathbf{x}$ lies on the geodesic between $\mathbf{y}$ and q. Therefore, the largest possible volume is $W(\alpha_M, \alpha_s, \arcsin(\hat{\alpha}_s + 2\varepsilon\delta))$, which corresponds to W_1 on Figure 5. It remains to show that W_1 is negligible compared with $W_1 + W_2$.

If $M < \sqrt{2}s$, then the main term of $W(\alpha_M, \alpha_s, \arcsin\left(\hat{\alpha}_s + \varepsilon\delta\right))$ is $\hat{\gamma}^d$ with

$$\hat{\gamma}^2 = \frac{2\left(M^2 s^2 + s^2(s+\varepsilon)^2 + M^2(s+\varepsilon)^2\right) - \left(M^4 + s^4 + (s+\varepsilon)^4\right)}{4(s+\varepsilon)^2}$$
$$= \frac{M^2+s^2}{2} - \frac{\left(M^2-s^2\right)^2}{4(s+\varepsilon)^2} - \frac{(s+\varepsilon)^2}{4}\,.$$

The main term of $W(\alpha_M, \alpha_s, \arcsin(\hat{\alpha}_s + 2\varepsilon\delta))$ is $\hat{\gamma}_1^d$ with

$$\hat{\gamma}_1^2 = \frac{M^2+s^2}{2} - \frac{\left(M^2-s^2\right)^2}{4(s+2\varepsilon)^2} - \frac{(s+2\varepsilon)^2}{4}\,.$$

It is easy to see that $M < \sqrt{2}s$ implies that $\hat{\gamma}_1^2 < \hat{\gamma}^2$. As a result, $W(\alpha_M, \alpha_s, \arcsin(\hat{\alpha}_s + 2\varepsilon\delta)) = o\left(W(\alpha_M, \alpha_s, \arcsin(\hat{\alpha}_s + \varepsilon\delta))\right)$ (similarly to the other proofs, it is easy to show that the effect of the other terms in Lemma 2 is negligible compared to $\left(\hat{\gamma}_1/\hat{\gamma}\right)^d$). Moreover, since at each step we reduce the radius of a spherical cap centered at q by at least $\varepsilon\delta$, any cap encountered in one iteration intersects with only a constant number of other caps, so their overall effect is negligible, which completes the proof.

Finally, let us note that as soon as we have $M > \sqrt{2}s$ (with $s > 1$), with probability $1 - o(1)$ we find the nearest neighbor in one step. $\square$

Having this result on "almost independence" of the consecutive steps, we can say that the overall success probability is $1-O\left(2^{\omega}e^{-S^{d+o(1)}}\right)$. Assuming $d \gg \log\log n$, we get $O\left(2^{\omega}e^{-S^{d(1+o(1))}}\right) = O\left(2^{\frac{\log n}{d}}e^{-\log n}\right) = o(1)$. This concludes the proof for the success probability $1-o(1)$ up to choosing suitable values for s and M.

Let us discuss the time complexity. With probability $1-o(1)$ the number of steps is $\Theta\left(\delta^{-1}\right)$: the upper bound was already discussed; the lower bound follows from the fact that $M\delta$ is the radius of a spherical cap, so we cannot make steps longer than $\arcsin(M\delta)$, and with probability $1-o(1)$ we start from a constant distance from q. The complexity of each step is $\Theta\left(f\cdot d\right) = \Theta\left(d^{1/2}\cdot M^{d}\right)$, so overall we get $\Theta\left(d^{1/2}\cdot 2^{\omega}\cdot M^{d}\right)$.

It remains to find suitable values for s and M. Before we continue, let us analyze the conditions under which we find exactly the nearest neighbor at the next step of the algorithm. Assume that the radius of a cap centered at q and covering the currently considered element is $\hat{\alpha}_s$ and $\hat{\alpha}_s = s\,\delta$, $\hat{\alpha}_M = M\delta$. Further assume that the radius of a spherical cap covering points at distance at most R from q is $\hat{\alpha}_r = r\delta = \sin R$ for some $r<1$. The following lemma gives the conditions for M and s such that at the next step of the algorithm we find the nearest neighbor $\bar{\mathbf{x}}$ with probability $1-o(1)$ given that $\bar{\mathbf{x}}$ is uniformly distributed within a distance R from q.

Lemma 11. *If for constant M, s, r we have $M^2 > s^2 + r^2$, then*

$$C(\alpha_r) - W(\alpha_M, \alpha_r, \arccos\alpha_s) \le C(\alpha_r)\beta^d$$

with some $\beta < 1$.

Proof. First, recall the lower bound for $C(\alpha_r)$: $C(\alpha_r) \ge \Theta\left(d^{-1/2}\right)\delta^d r^d$.

Since we have $M^2 > s^2 + r^2$, then asymptotically we have $\alpha_M < \alpha_s\alpha_r$, so the cases (1)-(2) of Lemma 2 should be applied (see Figure 2b). Let us estimate $\hat{\gamma}^2$ (Lemma 7):

$$\hat{\gamma}^2 \sim \delta^2\cdot\frac{2\left(M^2s^2+M^2r^2+s^2r^2\right)-\left(M^4+r^4+s^4\right)}{4\,s^2}$$
$$= \delta^2\cdot\frac{-\left(M^2-s^2-r^2\right)^2+4\,s^2r^2}{4\,s^2} = \delta^2 r^2\left(1-\Theta(1)\right).$$

Therefore $\hat{\gamma}^d \le \delta^d r^d \beta^d$ with some $\beta < 1$.

It only remains to estimate the other terms in the upper bound from Lemma 2:

$$\frac{\hat{\gamma}\,\alpha_r\,\hat{\alpha}_s}{\gamma|\alpha_M-\alpha_r\,\alpha_s|}\Theta\left(d^{-1}\right)\min\left\{d^{1/2},\frac{1}{\gamma}\right\} = O\left(d^{-1}\right),$$

from which the lemma follows. ∎

Now we are ready to finalize the proof. We solve c, R-ANN if either we have $\arcsin(s\delta) \le cR$ or we are sufficiently close to q to find the exact nearest neighbor $\bar{\mathbf{x}}$ in the next step of the algorithm. Let us analyze the first possibility. Let $\sin R = r\delta$, then we need $s < c\,r$. According to Lemma 9, it is sufficient to have $r\,c > 1$ and either $M \ge \sqrt{2}\,rc$ or $M^2 - \frac{M^4}{4r^2c^2} > 1$. Alternatively, according to Lemma 11, to find the exact nearest neighbor with probability $1-o(1)$, it is sufficient to reach such s that $M^2 > s^2 + r^2$. For this, according to Lemma 9, it is sufficient to have $s > 1$, $M^2 > s^2 + r^2$, and either $M \ge \sqrt{2}s$ or $M^2 - \frac{M^4}{4s^2} > 1$.

One can show that if the following conditions on M and r are satisfied, then we can choose an appropriate s for the two cases discussed above:

(a) $r\,c > 1$ and $M^2 > 2\,r^2\,c^2\left(1-\sqrt{1-\frac{1}{r^2c^2}}\right)$;

(b) $M^2 > \frac{2}{3}\left(r^2+1+\sqrt{r^4-r^2+1}\right)$.

To succeed, we need either (a) or (b) to be satisfied. The bound in (a) decreases with r ($r > 1/c$) and for $r = \frac{1}{c}$ it equals $\sqrt{2}$. The bound in (b) increases with r and for $r = 1$ it equals $\sqrt{2}$. To find a general bound holding for all r, we take the "hardest" $r \in (\frac{1}{c}, 1)$, where the bounds in (a) and (b) are equal to each other. This value is $r = \sqrt{\frac{4c^2}{(c^2+1)(3c^2-1)}}$ and it gives the bound $M > \sqrt{\frac{4c^2}{3c^2-1}}$ stated in the theorem.

B.4 Larger neighborhood in dense regime

As discussed in the main text, it could potentially be possible that taking $M = M(n) \gg 1$ improves the query time. The following theorem shows that this is not the case.

Theorem 1. *Let $M = M(n) \gg 1$. Then, with probability $1 - o(1)$, graph-based NNS finds the exact nearest neighbor in one iteration; time complexity is $\Omega\left(d^{1/2} \cdot 2^{\omega} \cdot M^{d-1}\right)$; space complexity is $\Theta\left(n \cdot d^{-1/2} \cdot M^d \cdot \log n\right)$.*

As a result, when $M \to \infty$, both time and space complexities become larger compared with constant M (see Theorem 1 from the main text).

Proof. When M grows with n, it follows from the previous reasoning that the algorithm succeeds with probability $1 - o(1)$. The analysis of the space complexity is the same as for constant M, so we get $\Theta\left(d^{-1/2} \cdot M^d \cdot n \cdot \log n\right)$. When analyzing the time complexity, we note that the one-step complexity is $\Theta\left(d^{-1/2} \cdot M^d\right)$. It is easy to see that we cannot make steps longer than $O\left(M \cdot 2^{-\omega}\right)$. This leads to the time complexity $\Omega\left(d^{1/2} \cdot M^{d-1} \cdot 2^{\omega}\right)$. $\square$

B.5 Proof of Theorem 2 (greedy search in sparse regime)

For sparse datasets, instead of radii, we operate with heights of spherical caps. In this case, we have $C(\alpha_1) \sim \frac{1}{n}$, i.e., $\alpha_1^2 \sim 1 - n^{-\frac{2}{d}} = 1 - 2^{-\frac{2}{\omega}} \sim \frac{2\ln 2}{\omega}$. We further denote $\frac{2\ln 2}{\omega}$ by δ.

We construct $G(M)$ using spherical caps with height α_M, $\alpha_M^2 = M\,\delta$, where M is some constant. Then, from Lemma 1, we get that the expected number of neighbors of a node is $f = \Theta\left(n\, d^{O(1)}(1 - M\delta)^{d/2}\right) = n^{1-M+o(1)}$. From this and Section B.2.2 the stated space complexity follows. The one-step time complexity $n^{1-M+o(1)}$ follows from Section B.2.1.

Our aim is to solve c, R-ANN with some $c > 1$, $R > 0$. If $R \geq \pi/2c$, then we can easily find the required near neighbor within a distance $\pi/2$, since we start $G(M)$-based NNS from such point. Let us consider any $R < \frac{\pi}{2c}$. It is clear that in this case we have to find the nearest neighbor itself, since $R\,c$ is smaller than the distance to the (non-planted) nearest neighbor with probability $1 + o(1)$. Note that $\alpha_c = \cos\frac{\pi}{2c} < \alpha_R := \cos R$.

Lemma 12. *Assume that $\alpha_M^2 = M\delta$, $\alpha_s^2 = s\delta$, $\alpha_\varepsilon^2 = \varepsilon\delta$, $M, s > 0$ are constants, and $\varepsilon \geq 0$ is bounded by a constant. If $s + M < 1$, then*

$$W(\alpha_M, \alpha_s, \arccos\alpha_\varepsilon) \geq \frac{1}{n} \cdot n^{\Omega(1)}\,.$$

Proof. Asymptotically, we have $\alpha_M > \alpha_s\alpha_\varepsilon$ and $\alpha_s > \alpha_M\alpha_\varepsilon$, so we are under the condition (3) in Lemma 2. First, consider the main term of $W(\alpha_M, \alpha_s, \arccos\alpha_\varepsilon)$:

$$\hat{\gamma}^d = \left(1 - \frac{M\delta + s\delta - 2\sqrt{M s\,\varepsilon\,\delta}\,\delta}{1 - \varepsilon\delta}\right)^{d/2} = e^{-\frac{d}{2}\delta\left(M+s+O\left(\sqrt{\delta}\right)\right)} = n^{-\left(M+s+O\left(\sqrt{\delta}\right)\right)} = \frac{1}{n} \cdot n^{\Omega(1)}.$$

It remains to multiply this by $\Theta\left(d^{-1}\right)$ and $C_l = C_{l,\alpha} + C_{l,\beta}$ (see Lemma 2). It is easy to see that $C_l = \Omega(1)$ in this case, so both terms can be included to $n^{\Omega(1)}$, which concludes the proof. $\square$

It follows from Lemma 12 that if $M + s < 1$, then we can reach a spherical cap with height $\alpha_s = \sqrt{s\delta}$ centered at q in just one step (starting from a distance at most $\pi/2$). And we get $g(n) = n^{\Omega(1)}$.

Recall that $M < \frac{\alpha_c^2}{\alpha_c^2+1}$ and let us take $s = \frac{1}{\alpha_c^2+1}$, then we have $M + s < 1$. The following lemma discusses the conditions for M and s such that at the next step of the algorithm we find $\bar{\mathbf{x}}$ with probability $1 - o(1)$.

Lemma 13. *If for constant M and s we have $M < s\alpha_R^2$, then*

$$C(\alpha_R) - W(\alpha_M, \alpha_R, \arccos\alpha_s) = C(\alpha_R)n^{-\Omega(1)}\,.$$

Proof. First, recall the lower bound for $C(\alpha_R)$: $C(\alpha_R) \ge \Theta\left(d^{-1/2}\right)(1-\alpha_R^2)^{d/2}$.

Note that since we have $M < s\alpha_R^2$, then the cases (1)-(2) of Lemma 2 should be applied (see Figure 2b). Let us estimate γ:

$$\gamma^2 = \frac{M\delta + \alpha_R^2 - 2\sqrt{Ms}\alpha_R\delta}{1-s\delta} = \alpha_R^2 + \delta\left(M - 2\sqrt{Ms}\alpha_R + s\alpha_R^2\right) + O\left(\delta^2\right)$$
$$= \alpha_R^2 + \delta\left(\sqrt{M} - \sqrt{s}\alpha_R\right)^2 + O\left(\delta^2\right) = \alpha_R^2 + \Theta(\delta)\,.$$

Therefore,

$$\hat\gamma^d = \left(1-\alpha_R^2\right)^{d/2}\left(1-\Theta(\delta)\right)^{d/2} = \left(1-\alpha_R^2\right)^{d/2} n^{-\Omega(1)}\,.$$

It only remains to estimate the other terms in the upper bound from Lemma 2:

$$\frac{\hat\gamma\,\alpha_R\,\hat\alpha_s}{\gamma|\alpha_M - \alpha_R\,\alpha_s|}\Theta\left(d^{-1}\right)\min\left\{d^{1/2}, \frac{1}{\gamma}\right\} = O\left(\delta^{-1}d^{-1}\right)\,,$$

from which the lemma follows. □

Note that in our case we have $M < \frac{\alpha_c^2}{\alpha_c^2+1} < \frac{\alpha_R^2}{\alpha_c^2+1} = s\alpha_R^2$. From this Theorem 2 follows.

C Long-range links

C.1 Random edges

The simplest way to obtain a graph with a small diameter from a given graph is to connect each node to a few random neighbors. This idea is proposed in [8] and gives $O(\log n)$ diameter for the so-called "small-world model". It was later confirmed that adding a little randomness to a connected graph makes the diameter small [2]. However, we emphasize that having a logarithmic diameter does not guarantee a logarithmic number of steps in graph-based NNS, since these steps, while being greedy in the underlying metric space, may not be optimal on a graph.

To demonstrate the effect of long edges, assume that there is a graph G', where each node is connected to several random neighbors by directed edges. For simplicity or reasonings, assume that we first perform NNS on G' and then continue on the standard NN graph G. It is easy to see that during NNS on G', the neighbors considered at each step are just randomly sampled nodes, we choose the one closest to q and continue the process, and all such steps are independent. Therefore, the overall procedure is basically equivalent to a random sampling of a certain number of nodes and then choosing the one closest to q (from which we then start the standard NNS on G).

Theorem 2. *Under the conditions of Theorem 1 in the main text, performing a random sampling of some number of nodes and choosing the one closest to q as a starting point for graph-based NNS does not allow to get time complexity better than $\Omega\left(d^{1/2}\cdot e^{\omega(1+o(1))}\cdot M^d\right)$.*

Proof. Assume that we sample $e^{l\omega}$ nodes with an arbitrary $l = l(n)$. Then, with probability $1-o(1)$, the closest one among them lies at a distance $\Theta\left(e^{-\frac{l\omega}{d}}\right)$. As a result, the overall time complexity becomes $\Theta\left(e^{-\frac{l\omega}{d}}\cdot d^{1/2}\cdot e^{\omega}\cdot M^d + d\cdot e^{l\omega}\right)$. If $l = \Omega(d)$, then the term $d\cdot e^{l\omega} = d\cdot e^{\Omega(d)\omega}$ dominates $d^{1/2}\cdot e^{\omega}\cdot M^d$, otherwise we get $\Theta\left(d^{1/2}\cdot e^{\omega(1+o(1))}\cdot M^d\right)$, which proves Theorem 2. □

C.2 Proof of Theorem 3 (effect of proper long edges)

Recall that we assume the following probability distribution:

$$\mathrm{P}(\text{edge from } u \text{ to } v) = \frac{\rho(u,v)^{-d}}{\sum_{w \neq u} \rho(u,w)^{-d}} \,. \tag{4}$$

First, we estimate the denominator. In the lemma below we consider only the elements w with $\rho(u,w) > n^{-\frac{1}{d}}$. However, it easily follows from the proof that adding only edges with $\rho(u,v) > n^{-\frac{1}{d}}$ does not affect the reasoning.

From Theorem 1 in the main text, we know that without long edges we need $O(n^{\frac{1}{d}})$ steps, which is less than $\log^2 n$ for $d > \frac{\log n}{2 \log \log n}$. So, in this case Theorem 3 follows from Theorem 2. Hence, in the lemma below we can assume that $d < \frac{\log n}{2 \log \log n}$.

Lemma 14. *If $d < \frac{\log n}{2 \log \log n}$, then*

$$\mathbb{E}\left(\rho(u,w)^{-d}\right) = \Theta\left(\frac{\log n}{\sqrt{d}}\right) .$$

Proof. Note that $\mathbb{E}\left(\rho(u,w)^{-d}\right) = \mathbb{E}\rho^{-d}(\mathbf{1},w)$, where $\mathbf{1} = (1,0,\ldots,0)$. So, similarly to Lemma 1,

$$\mathbb{E}\left(\rho(\mathbf{1},w)^{-d}\right) = \frac{\mu(S^{d-1})}{\mu(S^d)} \int\limits_{-1}^{\cos(n^{-\frac{1}{d}})} \left(1-x^2\right)^{\frac{d-2}{2}} (\arccos x)^{-d} dx \,.$$

From Stirling's approximation, we have $\frac{\mu(S^{d-1})}{\mu(S^d)} = \Theta(\sqrt{d})$. After replacing $y = \arccos x$, the integral becomes

$$\int\limits_{n^{-\frac{1}{d}}}^{\pi} y^{-d} \sin^{d-1}(y) dy = \int\limits_{n^{-\frac{1}{d}}}^{\pi} \frac{1}{y}\left(\frac{\sin(y)}{y}\right)^{d-1} dy < \int\limits_{n^{-\frac{1}{d}}}^{\pi} \frac{1}{y} dy = \Theta\left(\frac{\ln n}{d}\right) .$$

On the other hand, for $d < \frac{\log n}{2 \log \log n}$:

$$\mathbb{E}\left(\rho(\mathbf{1},w)^{-d}\right) = \Theta(\sqrt{d}) \int\limits_{n^{-\frac{1}{d}}}^{\pi} \frac{1}{y}\left(\frac{\sin(y)}{y}\right)^{d-1} dy > \Theta(\sqrt{d}) \int\limits_{n^{-\frac{1}{d}}}^{\frac{1}{\sqrt{d}}} \frac{1}{y}\left(\frac{\sin(y)}{y}\right)^{d-1} dy.$$

Since on this interval we have $\left(\frac{\sin(y)}{y}\right)^{d-1} = \Theta(1)$, we can continue:

$$\mathbb{E}\left(\rho(\mathbf{1},w)^{-d}\right) > \Theta(\sqrt{d}) \int\limits_{n^{-\frac{1}{d}}}^{\frac{1}{\sqrt{d}}} \frac{1}{y} dy = \Theta(\sqrt{d}) \left(\frac{\ln n}{d} - \frac{1}{2}\ln d\right) = \Theta\left(\frac{\ln n}{\sqrt{d}}\right) .$$

As a result, we get $\mathbb{E}\rho(\mathbf{1},w)^{-d} = \Theta\left(\frac{\log n}{\sqrt{d}}\right)$. □

Also, from the proof above it follows that $\mathbb{E}\rho(u,w)^{-2d} < \Theta(\sqrt{d}) \int\limits_{n^{-\frac{1}{d}}}^{\pi} \frac{1}{y^{d+1}} dy = O(\frac{n}{\sqrt{d}})$. Let

$$\mathrm{Den} = \sum_{w : \rho(u,w) > n^{-\frac{1}{d}}} \rho(u,w)^{-d},$$

so $\mathbb{E}\,\mathrm{Den} = \Theta\left(\frac{n\ \log n}{\sqrt{d}}\right)$ and

$$\mathbb{E}\,\mathrm{Den}^2 = n\,\mathbb{E}\rho(u,w)^{-2d} + n(n-1)\left(\mathbb{E}\rho(u,w)^{-d}\right)^2 = O\left(\frac{n^2}{\sqrt{d}} + \frac{n^2\ln^2 n}{d} - \frac{n\ln^2 n}{d}\right).$$

Finally, from Chebyshev's inequality, we get

$$\mathrm{P}\left(|\mathrm{Den} - \mathbb{E}\mathrm{Den}| > \frac{\mathbb{E}\mathrm{Den}}{2}\right) \le \frac{4\,\mathrm{Var}(\mathrm{Den})}{(\mathbb{E}\mathrm{Den})^2} = O\left(\frac{\sqrt{d}}{\log^2 n}\right) = o(1).$$

So, we may further replace the denominator of Equation (4) by $O\left(\frac{n\ \log n}{\sqrt{d}}\right)$.[2]

We are ready to prove the theorem. We split the search process on a sphere into $\log n$ phases and show that each phase requires $O(\log n)$ steps. Phase j consists of the following nodes: $\{u : t_{j+1} < \rho(u,q) \le t_j\}$, where $t_j = \frac{\pi}{2}\ t^j = \frac{\pi}{2}\ \left(1 - \frac{1}{d}\right)^j$.

We start at a distance at most $\frac{\pi}{2}$, this corresponds to $j = 0$. Recall that the nearest neighbor (in the dense regime) is at a distance about $2^{-\frac{\log n}{d}}$. Then, the number of phases needed to reach the nearest neighbor is

$$k \sim -\frac{1}{\log\left(1 - \frac{1}{d}\right)} \cdot \frac{\log n}{d} \sim \log n\,.$$

Suppose we are at some node belonging to a phase j. Let us prove the following inequality for the probability of making a step to a phase with a larger number:

$$\mathrm{P}(\text{make a step to a closer phase}) > \frac{\Theta(1)}{\log n}\,.$$

In the polar coordinates, we can express this probability as

$$\begin{aligned}\mathrm{P}(\text{make a step to a closer phase}) &= \frac{(d-1)\sqrt{d}}{2\pi\ \log n}\int_{\cos t_{i+1}}^{1}\int_{-\arccos\frac{\cos(t_{i+1})}{r}}^{\arccos\frac{\cos(t_{i+1})}{r}}\left(\sqrt{1-r^2}\right)^{d-3} \\ &\quad\cdot\left(\arccos(\sin(t_i)r\sin(\phi) + \cos(t_i)r\cos(\phi))\right)^{-d} r\,d\phi\,dr \\ &= \frac{\Theta(d^{\frac{3}{2}})}{\log n}\int_{\cos t_{i+1}}^{1}\int_{-\arccos\frac{\cos(t_{i+1})}{r}}^{\arccos\frac{\cos(t_{i+1})}{r}}\left(\sqrt{1-r^2}\right)^{d-3}\left(\arccos(r\cos(t_i - \phi))\right)^{-d} r\,d\phi\,dr.\end{aligned}$$

Let $r = \cos(\psi)$ and $\phi = t_j - \phi$, then the integral becomes

$$\int_0^{t_{j+1}} \cos\psi\,(\sin\psi)^{d-2}\int_{t_j - \arccos\frac{\cos(t_{j+1})}{\cos\psi}}^{t_j + \arccos\frac{\cos(t_{j+1})}{\cos\psi}} (\arccos(\cos\psi\cos(\phi))^{-d} d\phi d\psi\,.$$

From convexity of $\log\cos\sqrt{x}$, it follows that $\forall\psi,\phi \in [0,\frac{\pi}{2}]$ we have

$$\arccos(\cos\psi\cos\phi) \le \sqrt{\psi^2 + \phi^2}\,,$$

$$\arccos\frac{\cos(t_{j+1})}{\cos\psi} \ge \sqrt{t_{j+1}^2 - \psi^2}\,,$$

$$\sin(\psi) \ge \psi - \frac{\psi^3}{6}\,.$$

[2]More formally, our analysis below is conditioned on the fact that the denominator is less than $\frac{C\,n\log n}{\sqrt{d}}$ for some constant $C > 0$. The probability that it does not hold is $o(1)$ and for such nodes we can just assume that we do not use long edges.

We use these bounds since we need a lower bound for the integral. Also, we replace the upper limit of the inner integral with t_j and consider $\psi = t_j\psi$ and $\phi = t_j\phi$:

$$\int\limits_0^t F(t_j,\psi)\psi^{d-2} \int\limits_{1-\sqrt{t^2-\psi^2}}^{1} (\sqrt{\psi^2+\phi^2})^{-d}\, d\phi\, d\psi\,,$$

where $F(t_j,\psi) = \cos(t_j\ \psi)\left(1-\frac{t_j^2\psi^2}{6}\right)^{d-2}$.

Consider the inner integral:

$$\int\limits_{1-\sqrt{t^2-\psi^2}}^{1} (\sqrt{\psi^2+\phi^2})^{-d}\,\frac{1}{2\phi}d\phi^2 > \frac{1}{2}\int\limits_{1-2\sqrt{t^2-\psi^2}+t^2-\psi^2}^{1} (\psi^2+x)^{-\frac{d}{2}}\,dx$$
$$= \frac{1}{d-2}\left((1-2\sqrt{t^2-\psi^2}+t^2)^{-\frac{d-2}{2}} - (1+\psi^2)^{-\frac{d-2}{2}}\right).$$

Substitute the second term to the original integral and estimate it from above:

$$\int_0^t F(t_j,\psi)\psi^{d-2}(1+\psi^2)^{-\frac{d-2}{2}}\,d\psi \le \int_0^t \left(\frac{\psi^2}{1+\psi^2}\right)^{\frac{d-2}{2}}\,d\psi = o\left(\frac{1}{\sqrt{d}}\right).$$

Now we estimate from below the second term $\int\limits_0^t \left(\frac{\psi^2}{1-2\sqrt{t^2-\psi^2}+t^2}\right)^{\frac{d-2}{2}}\,d\psi$.

Note that if $\psi = \frac{2}{\sqrt{d}}$ and $t = 1-\frac{1}{d}$, then

$$\left(\frac{\psi^2}{1-2\sqrt{t^2-\psi^2}+t^2}\right)^{\frac{d-2}{2}} = \left(\frac{\frac{4}{d}}{1-2\sqrt{1-\frac{2}{d}+\frac{1}{d^2}-\frac{4}{d}}+1-\frac{2}{d}+\frac{1}{d^2}}\right)^{\frac{d-2}{2}}$$
$$> \left(\frac{\frac{4}{d}}{\frac{4}{d}+\frac{11}{d^2}+\frac{1}{d^2}}\right)^{\frac{d-2}{2}} = e^{-\frac{3}{2}}(1+o(1))\,.$$

Similarly, it can be shown that if $\psi = \frac{3}{\sqrt{d}}$, then

$$\left(\frac{\psi^2}{1-2\sqrt{t^2-\psi^2}+t^2}\right)^{\frac{d-2}{2}} > \Theta(1).$$

So, for $\psi \in [\frac{2}{\sqrt{d}}, \frac{3}{\sqrt{d}}]$ this fraction is greater than some constant (as well as $F(t_j,\psi)$) and the derivative does not change the sign on this segment. As a result,

$$\int_0^t F(t_j,\psi)\left(\frac{\psi^2}{1-2\sqrt{t^2-\psi^2}+t^2}\right)^{\frac{d-2}{2}}\,d\psi > \frac{\Theta(1)}{\sqrt{d}}\,.$$

And finally,

$$\mathrm{P}(\text{make a step in to a closer phase}) > \frac{\Theta(1)}{\log n}\,.$$

To sum up, there are $O(\log n)$ phases and the number of steps in each phase is geometrically distributed with the expected value $O(\log n)$. From this the theorem follows.

C.3 Proof of Corollary 2

We have

$$\mathrm{P}(\text{short-cut step within } \log n \text{ trying}) = 1 - (1-P)^{\log n} = \left(1 - \frac{1}{e}\right)(1 - o(1)), \tag{5}$$

where P is the probability corresponding to one long-range edge, which is estimated in the proof above.

Also, since $d \gg \log\log n$, we have $\frac{M^d}{\sqrt{d}} > \log n$, so the step complexity is the same.

It is easy to see that increasing the number of shortcut edges does not improve the asymptotic complexity, since the probability in (5) is already constant.

C.4 Proof of Lemma 1 (effect of pre-sampling)

For convenience, in this proof we assume that the overall number of elements is $n+1$ instead of n which does not affect the analysis.

Let v be the k-th neighbor for the source node u. For the initial distribution we have:

$$\mathrm{P}(\text{edge from } u \text{ to } v) \sim \frac{1}{k \ln n}\,.$$

By pre-sampling of n^{φ} nodes, we modify this probability to

$$\mathrm{P}(\text{edge from } u \text{ to } v | v \text{ is sampled}) \cdot \mathrm{P}(v \text{ is sampled}). \tag{6}$$

The second term above is equal to $\frac{n^{\varphi}}{n}$. Assuming that $k = n^{\alpha} > n^{1-\varphi}$, we can estimate the probability above. Below by l we denote the rank of v in the selected subset and obtain:

$$\begin{aligned}
&\mathrm{P}(\text{edge from } u \text{ to } v | v \text{ is sampled}) \cdot \mathrm{P}(v \text{ is sampled}) \\
&\quad = \frac{n^{\varphi}}{n} \sum_{l=1}^{\min(n^{\alpha}, n^{\varphi})} \binom{n^{\varphi}-1}{l-1} \left(\frac{n^{\alpha}-1}{n-1}\right)^{l-1} \left(\frac{n-n^{\alpha}}{n-1}\right)^{n^{\varphi}-l} \frac{1}{l}\, \frac{1}{\ln n^{\varphi}} \\
&\quad = \frac{1}{\varphi\, n\, \ln n} \sum_{l=1}^{\min(n^{\alpha}, n^{\varphi})} \binom{n^{\varphi}}{l} \left(\frac{n^{\alpha}-1}{n-1}\right)^{l-1} \left(\frac{n-n^{\alpha}}{n-1}\right)^{n^{\varphi}-l} \\
&\quad = \frac{\Theta(1)}{\varphi\, n^{\alpha} \ln n} \sum_{l=1}^{\min(n^{\alpha}, n^{\varphi})} \binom{n^{\varphi}}{l} \left(\frac{n^{\alpha}-1}{n-1}\right)^{l} \left(\frac{n-n^{\alpha}}{n-1}\right)^{n^{\varphi}-l}\,.
\end{aligned}$$

Let us analyze the sum above. First, it is easy to see that it is less than 1. Second, if $n^{\varphi} \le n^{\alpha}$, then the sum is "almost equal" to 1 (without one term corresponding to $l = 0$, which we analyze below). Otherwise, we know that for a binomial distribution its median cannot lie too far away from the mean (see, e.g., [3]). Since $\alpha > 1 - \varphi$, we have

$$\min(n^{\alpha}, n^{\varphi}) \ge \frac{n^{\varphi}(n^{\alpha}-1)}{n-1} + 1 = \mathbb{E}\,\mathrm{Bin}(n^{\varphi}, \frac{n^{\alpha}-1}{n-1}) + 1 > \mathrm{median}(\varphi, \alpha).$$

Hence,

$$\sum_{l=0}^{\min(n^{\alpha}, n^{\varphi})} \binom{n^{\varphi}}{l} \left(\frac{n^{\alpha}-1}{n-1}\right)^{l} \left(\frac{n-n^{\alpha}}{n-1}\right)^{n^{\varphi}-l} > \frac{1}{2}.$$

Note that we added one term corresponding to $l = 0$, but it is easy to see that in the worst case it is about $\frac{1}{e}$. Namely, for $l = 0$:

$$\binom{n^{\varphi}}{0} \left(\frac{n^{\alpha}-1}{n-1}\right)^{0} \left(\frac{n-n^{\alpha}}{n-1}\right)^{n^{\varphi}-0} = \left(\frac{n-n^{\alpha}}{n-1}\right)^{n^{\varphi}} < \left(1 - \frac{n^{1-\varphi}-1}{n-1}\right)^{n^{\varphi}} = \frac{1}{e}(1 + o(1)).$$

D Proof of Theorem 4 (effect of beam search)

Let us call a spherical cap of radius $L\delta$ centered at x an L-neighborhood of x. We first show that the subgraph of $G(M)$ induced by the L-neighborhood contains a path from a given element to the nearest neighbor of the query with high probability.

For random geometric graphs in d-dimensional Euclidean space (for fixed d) it is known that the absence of isolated nodes implies connectivity [6, 7]. However, generalizing [6, 7] to our setting is non-trivial, especially taking into account that the dimension grows as the logarithm of the number of elements in the L-neighborhood. In our case, it is easy to show that with high probability there are no isolated nodes. Moreover, the expected degree is about S^d for some $S > 1$. Hence, it is possible to prove that the graph is connected. However, for simplicity, we prove a weaker result: for two fixed points, there is a path between them with high probability.

Let us denote by N the number of nodes in the L-neighborhood. According to Lemma 3, with high probability, this value is $\Theta\left(d^{-1/2}L^d\right)$. So, we further assume that there are $N = \Theta\left(d^{-1/2}L^d\right)$ points uniformly distributed within the L-neighborhood.

Let us make the following observation that simplifies the reasoning. Consider the L-neighborhood of q. Let us project all N points to the boundary of a neighborhood (moving them along the rays starting at q) and construct a new graph on these elements using the same M-neighborhoods. It is easy to see that this operation may only remove some edges and never adds new ones. Therefore, it is sufficient to prove connectivity assuming that N nodes are uniformly distributed on a boundary of the L-neighborhood. This allows us to avoid boundary effects and simplify reasoning.

Let p_1 be the probability that two random nodes are connected. This probability is the volume of the M-neighborhood of a node normalized by the volume of the boundary of the L-neighborhood. Under the conditions on M and L, one can show that p_1 is at least $\left(\frac{S}{L}\right)^d$ for some constant $S > 1$.

We fix any pair of nodes u, v and estimate the probability that there is a path of length k between them. We assume that $k \to \infty$ and $k = o(\sqrt{Np_1})$, which is possible to achieve since $p_1 N \to \infty$. We show that the probability of not having such a path is $o(1)$.

Let us denote by $P_k(u, v)$ the number of paths of length k between u and v. Then,

$$\mathbb{E}P_k(u,v) \sim \binom{N-2}{k-1}(k-1)!p_1^k \gtrsim N^{k-1}\left(\frac{S}{L}\right)^{dk} = \left(\frac{N}{L^d}\right)^k \frac{1}{N} S^{dk} .$$

We have $\mathbb{E}P_k(u, v) \to \infty$ if $k \to \infty$.

To claim concentration near the expectation, we estimate the variance. Note that $P_k(u, v) = \mathbb{E}(\sum_i I_i)$, where I_i indicates the event that a particular path is present and i indexes all possible paths of length k. Then, we can estimate

$$\begin{aligned}
\mathbb{E}P_k(u,v)^2 - (\mathbb{E}P_k(u,v))^2 &= \mathbb{E}\left(\sum_i I_i\right)^2 - (\mathbb{E}P_k(u,v))^2 \\
&= \sum_i \mathrm{P}(I_i = 1) \sum_j \left(\mathrm{P}(I_j = 1 | I_i = 1) - \mathrm{P}(I_j = 1)\right) \\
&= \mathbb{E}P_k(u,v) \sum_j \left(\mathrm{P}(I_j = 1 | I_i = 1) - \mathrm{P}(I_j = 1)\right) .
\end{aligned}$$

It is easy to see that $\sum_j \left(\mathrm{P}(I_j = 1|I_i = 1) - \mathrm{P}(I_j = 1)\right) = o\left(\mathbb{E}P_k(u, v)\right)$. Indeed, for most pairs of paths we have $\mathrm{P}(I_j = 1|I_i = 1) \sim \mathrm{P}(I_j = 1) \sim p_1^k$ since they do not share any intermediate nodes. Let us show that the contribution of the remaining pairs is small. The fraction of pairs of paths sharing k_0 intermediate nodes is $O\left(\frac{k^{2k_0}}{N^{k_0}}\right)$. Then, $\mathrm{P}(I_j = 1|I_i = 1) \le \mathrm{P}(I_j = 1)/p_1^{k_0}$, since in the worst case the paths may share k_0 consecutive edges. Since $k^2 \ll Np_1$, the relative contribution is $\sum_{k_0 \ge 1} O\left(\left(\frac{k^2}{Np_1}\right)^{k_0}\right) = o(1)$. Therefore, we get $\mathrm{Var}(P_k(u, v)) = o\left(\left(\mathbb{E}P_k(u, v)\right)^2\right)$

Finally, it remains to apply Chebyshev's inequality and get that $\mathrm{P}(P_k(u, v) < \mathbb{E}P_k(u, v)/2) = o(1)$, so at least one such path exists with high probability.

Now we are ready to prove the theorem. Let us prove that $G(M)$-based NNS succeeds with probability $1 - o(1)$. It follows from Lemma 9 (and the discussion below it) that under the conditions on M and L, greedy $G(M)$-based NNS reaches the L-neighborhood of the query with probability $1 - o(1)$.

Thus, with probability $1 - o(1)$ we reach the L-neighborhood within which there is a path to the nearest neighbor. Recall that we assume beam search with $\frac{CL^d}{\sqrt{d}}$ candidates. Choosing large enough C, we can guarantee that the number of candidates is larger than the number of elements in the L-neighborhood. This implies that all reachable elements inside the L-neighborhood will finally be covered by the algorithm.

Finally, it remains to analyze the time complexity. To reach the L-neighborhood, we need $\Theta\left(d^{1/2} \cdot \log n \cdot M^d\right)$ operations (recall that the number of steps can be bounded by $\log n$ due to long edges). Then, to fully explore the L-neighborhood, we need $O\left(L^d \cdot M^d\right)$. For $d > \log\log n$ the first term is negligible compared to the second one, so the required complexity follows.

E Comparison with the results of [4]

Here we extend the related work from the main text and discuss in more detail how our research differs from the results of [4].

Laarhoven [4] analyzes time and space complexity for graph-based NNS in sparse regime when $d \gg \log n$. He considers plain NN graphs and allows multiple restarts. In contrast, we consider both regimes and assume only one iteration of a graph-based search. We do not consider multiple restarts since it is non-trivial to rigorously prove that restarts can be assumed "almost independent" (see Section A.3, proof of Lemma 15, [4]). As a result, for sparse datasets, we consider a slightly weaker setting with only one iteration, but all results are formally proven. Our result for sparse regime (Theorem 2 in the main text) corresponds to the case $\rho_q = \rho_s$ from [4].

Also, in Section A, we state new bounds for the volumes of spherical caps' intersections, which are needed for the rigorous analysis in both sparse and dense regimes. We could not use the results of [1] since parameters defining spherical caps are assumed to be constant there, while they can tend to 0 or 1 in dense and sparse regimes.

We also address the problem of possible dependence between consecutive steps of the algorithm (Lemma 10). While we prove that it can be neglected, it is important for rigorous analysis.

Most importantly, we analyze the dense regime and additional techniques (shortcuts and beam search), which are essential for the effective graph-based search. Interestingly, shortcut edges are useful only in the dense regime.

F Additional experiments

F.1 Dense vs sparse setting

Let us discuss our intuition on why real datasets are "more similar" to dense rather than sparse synthetic ones.

In the sparse regime, all elements are almost at the same distance from each other, and even in the moderate regime ($d \propto \log(n)$), the distance to the nearest neighbor must be close to a certain constant. In contrast, the dense regime implies high proximity of the nearest objects. While real datasets are always finite and asymptotic properties cannot be formally verified, we still can compare the properties of real and synthetic datasets. We plotted the distribution of the distance to the nearest neighbor (see Figure 6) and see that for the SIFT dataset, the obtained distribution is more similar to the ones in the dense regime. This is further supported by the literature which estimates the intrinsic dimension of real data. For example, for the SIFT dataset with 128-dimensional vectors, the estimated intrinsic dimension is 16 [5]. Thus, we conclude that the analysis of the dense regime is important.

F.2 Parameters of algorithms

In this section, we specify additional hyperparameters used in our experiments.

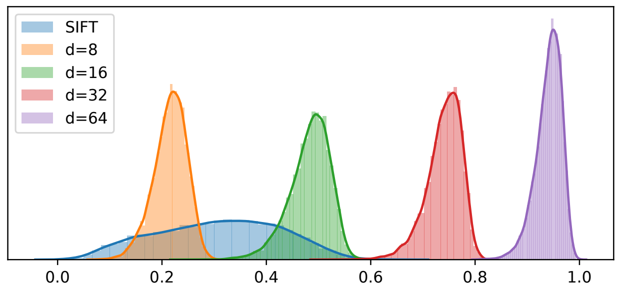


Figure 6: The distribution of the distance to the nearest neighbor for the SIFT dataset and synthetic uniform data for different dimensions and the same size (1M)

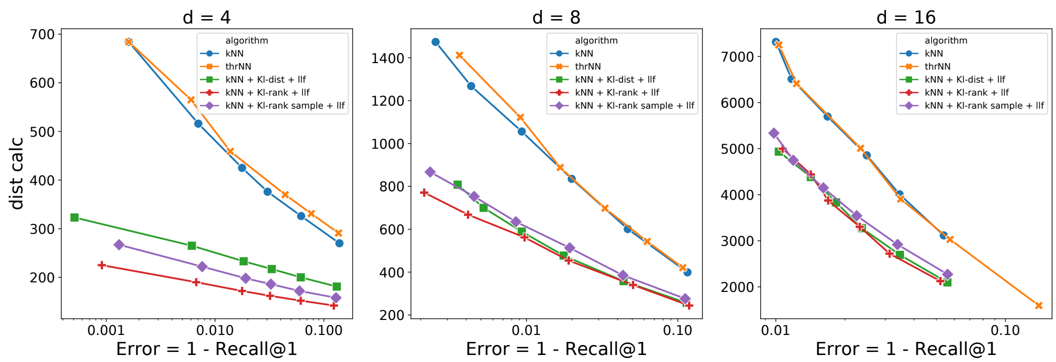


Figure 7: The effect of KNN and KL approximations

The number of edges used in KL, when is not explicitly specified, is equal to 15, which is close to $\ln n$.

The number of edges in KNN graphs is dynamic when the beam search is not used. When the beam search is used, the number of edges for synthetic datasets is 8 for $d = 2$, 10 for $d = 4$, 16 for $d = 8$, 20 for $d = 16$, and 25 for all real datasets.

The dimension we use for DIM-RED is 64 for GIST, 32 for SIFT, 48 for DEEP, 128 for GloVe.

F.3 Additional experimental results

In Figure 7, we show that several approximations discussed in the main text do not affect the quality of graph-based NNS significantly (in the uniform case). Namely,

- Connecting a node to other nodes at a distance smaller than some constant (THRNN) and to the fixed number of nearest neighbors (KNN) lead to graph-based algorithms with similar performance;
- Pre-sampling of $\sqrt{n}$ nodes when adding shortcut edges (SAMPLE) lowers the quality, but not substantially;
- Rank-based probabilities for shortcut edges (KL-RANK) can lead to even better quality than distance-based (KL-DIST).

In Figure 8, we illustrate how the number of long-range edges affects the quality of the algorithm. Let us note that 16 is close to $\log n$ discussed in Corollary 2 of the main text. Figure 8 shows that this value is indeed close to being optimal, especially for the high-accuracy regime, which is a focus of the current research. However, it also seems that the optimal number of long edges may depend on d: on Figure 8, the relative performance of graphs with 32 long edges is improving as d grows.

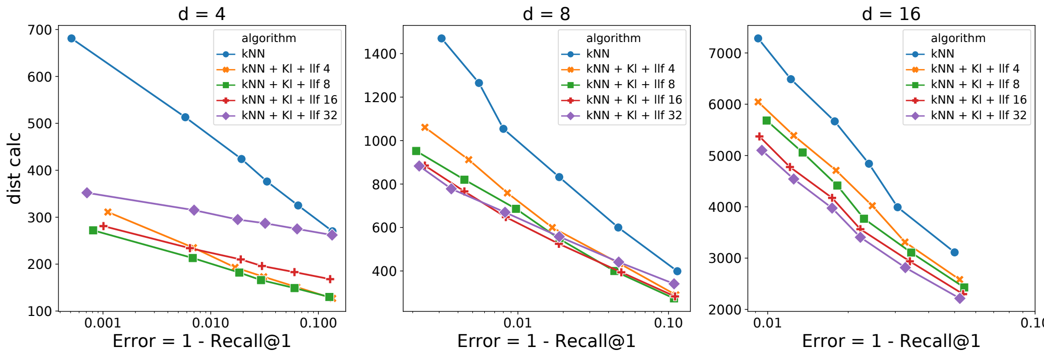


Figure 8: The effect of the number of long-range edges

References

[1] A. Becker, L. Ducas, N. Gama, and T. Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, pages 10–24, 2016.

[2] B. Bollobás and F. R. K. Chung. The diameter of a cycle plus a random matching. *SIAM Journal on discrete mathematics*, 1(3):328–333, 1988.

[3] K. Hamza. The smallest uniform upper bound on the distance between the mean and the median of the binomial and poisson distributions. *Statistics & Probability Letters*, 23(1):21–25, 1995.

[4] T. Laarhoven. Graph-based time-space trade-offs for approximate near neighbors. In *34th International Symposium on Computational Geometry (SoCG 2018)*, 2018.

[5] E. Levina and P. J. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in neural information processing systems*, pages 777–784, 2005.

[6] M. D. Penrose. On k-connectivity for a geometric random graph. *Random Structures & Algorithms*, 15(2):145–164, 1999.

[7] M. D. Penrose et al. Connectivity of soft random geometric graphs. *The Annals of Applied Probability*, 26(2):986–1028, 2016.

[8] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *Nature*, 393:440–442, 1998.