Theorem 0.1. *Given n i.i.d samples of a Lipschitz probability density on a d-dimensional manifold of bounded curvature, a kNN graph with $k = 2^d \cdot O(\log n)$ neighbors is a 1-spanner of the edge-squared graph with probabililty $> 1 - \frac{1}{n^{\Omega(1)}}$, for all large enough n. Here, the edge weights of the kNN graph are equal to the Euclidean distance squared.*

The big O in k hides a constant of approximately 4, if you want a 1-spanner with probability $> 1 - 1/n^2$. This means that, if intrinsic dimension d is small, such a spanner can be practically computed. This constant factor is independent of the Lipschitz constant or the bounds on the probability density. (The Lipschitz constant and the bounds on the probability density simply change how large n is before the claim kicks in.)

Proof. We prove it first for a uniform distribution on a square, and then claim (without proof) that the general bound follows. Let $d_2(x, y)$ be the edge squared distance between x and y, where x and y are vertices in the $i.i.d$ sample. It suffices to show that for any edge xy in which

$$||x - y||^2 = d_2(x, y)$$

is in the kNN graph with probability $\frac{1}{n^{\Omega(1)}}$. If this were true, then our result would follow by union bound.

Consider edge xy where xy is not in the kNN graph. Then the probability that $||x-y||^2 = d_2(x, y)$ is small. How small? Well, we know it is less than the probability that any sample point is in the ball with diameter xy.

Since xy is not in the kNN graph, we know that there are at least $p \geq k$ points in the ball centered at x with radius xy. The probability that none of these points is located in the ball with diameter xy is equal to:

$$\left(1 - \frac{\text{Volume-of-ball-with-diameter-xy}}{\text{Volume-of-ball-with-radius-xy-centered-at-x}}\right)^p$$

$$= \left(1 - \frac{1}{2^d}\right)^p$$

when k is set to be $2^d \cdot C \log n$, then the above expression is bounded above by:

$$\frac{1}{n^C}.$$

Taking the union bound over all such xy that are not in the kNN graph, says that with probability $< \frac{1}{n^{C-2}}$, the kNN graph is a 1-spanner of the edge-squared metric on n i.i.d points selected in the uniform sample.

□

Note: I screwed up here a bit: this only holds when the ball with diameter xy is strictly inside the distribution. This may be minor, or it may break my entire proof.