Response to Reviewer #1 We thank the reviewer for taking the time to review our submission and for their helpful suggestions. Regarding recovery of k-NN graphs, you are correct that a best-k strategy could be used to identify individual edges. The main challenge would be working out the condition for when the triangle inequality implies elimination. A point is eliminated if we can certify that there are k closer points due to either triangle inequalities or symmetry. That is ensured if for a given j, there are k distinct previous points i that satisfy the condition of Lemma 4.3. In the case of clustered data, as long as any cluster has at least k points, the same separation condition will apply, and this should give an additional factor of k in theorems 4.5 and 4.6. We will add a discussion of extending to k-NN graphs to the paper and correct the references to the algorithms.

Response to Reviewer #2 We thank the reviewer for their comments on our work and all helpful suggestions for how to improve it. To the best of our knowledge, we are unaware of any prior work that uses noisy triplet queries to learn the NN-graph. In the noiseless setting, there are techniques for nearest neighbor search using triplets [1,2], which potentially could be modified and used as a sub-routine by an algorithm that finds the NN-graph when there is no noise. We will elaborate on works relevant to the noiseless triplet setting in the related work section. Regarding provided code, we included example code zipped in the supplementary to preserve double blindness and apologize for the confusion with the checklist. On the note of the metric assumption, in the supplementary, we compare against an implementation that does not use the triangle inequality. Compared to the method that uses triangle inequality, we see slightly worse initial performance but similar gains over random sampling at higher accuracy levels. We will move this discussion to the main body to highlight it. Thank you also for pointing out the typos. We will correct them, simplify the notation, and define the quantities in Lemma 3.1 before the statement of the Lemma.

Response to Reviewer #3 We thank the reviewer for taking the time to review our work and all their suggestions, especially about how to clarify and contextualize the selection criteria.

Re: quantification of entropy over NN-graphs. That is an interesting way of looking at the problem. In some problem instances, the number of queries made by our algorithm is within a constant factor of the minimum bits required to specify the answer from a list of all possible NN-graphs. There are a total of n^{n-1} possible NN-graphs, hence each can be specified using $(n-1)\log n$ bits. If the dataset consists of hierarchical clusters as in the condition for Theorem 4.6, then `ANNEasy` finds the NN-graph after making $O(n \log(n)\overline{\Delta^{-2}})$ queries. Thus even though our noisy distance oracle is weaker than an oracle that answers arbitrary yes/no queries (e.g. membership queries of the form: is the true NN-graph present in a particular subset of all possible NN-graphs?), we are able to identify the NN-graph within a factor of the optimum number of queries and a multiplicative penalty to account for the noise in each answer.

The effect of having a weaker oracle can be seen in a different problem instance where $\Omega(n^2)$ distance queries are necessary to identify the NN-graph by any algorithm that uses the weaker oracle. For example, consider a dataset consisting of points $x_i = e_i + \epsilon \in \mathbb{R}^n \ \forall i$ where e_i is the unit vector with 1 in the ith component and 0 elsewhere, and ϵ is small independent zero-mean noise. Then all points are roughly at the same distance from each other, and for finding x_{i^*} we have to query $\mathsf{Q}(i,j)$ for all $j \neq i$. This is made more explicit in the discussion following Theorem 4.4.

Re: selection criteria for $\mathsf{Q}(i,j)$. Our algorithm iterates over points x_i in the dataset and finds x_{i^*} before starting the procedure for x_{i+1}. In the ith round, we use a modified successive elimination algorithm for bandit best-arm identification to find x_{i^*}. It is known that this algorithm matches instance-dependent lower bounds for best-arm identification within $\log$-factors [3]. In that sense, for a given x_i, our algorithm optimally selects x_j while querying $\mathsf{Q}(i,j)$ to find x_{i^*}. We will add this discussion before our reference to Algorithm 2. The triangle inequality bounds used for elimination are also optimal. [4] show that this Floyd-Warshall style approach yields the tightest upper and lower bounds on the distance matrix in the entry-wise L_1 norm.

The order in which the points $\{x_i\}$ are processed follows their subscript index, which is randomly chosen and fixed before starting the algorithm. Different orders in which $\{x_i\}$ are processed can affect the query complexity of our algorithm as discussed in the paragraph after Theorem 4.4. However it is not always possible to find an optimal order for $\{x_i\}$ from only noisy distance samples without assumptions on the metric. For example, if the oracle is noiseless, there are datasets where the pair (i,j) with the smallest $d_{i,j}$, must be queried within the first n queries to identify the NN-graph using the minimum number of queries. Since that cannot be ensured by an algorithm that only has access to information via a distance oracle, it is not possible to achieve the minimum number of queries in such examples.

[1] Haghiri, S., Ghoshdastidar, D., Luxburg, U.v.. (2017). *Comparison-Based Nearest Neighbor Search.* Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, in PMLR 54:851-859

[2] Houle, Michael E., and Michael Nett (2013), *Rank cover trees for nearest neighbor search*, International Conference on Similarity Search and Applications. Springer, Berlin, Heidelberg.

[3] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. 2016. *On the complexity of best-arm identification in multi-armed bandit models.* J. Mach. Learn. Res. 17, 1 (January 2016), 1-42.

[4] Singla, Adish, Sebastian Tschiatschek, and Andreas Krause. *Actively learning hemimetrics with applications to eliciting user preferences.* International Conference on Machine Learning. 2016.