

Supplementary Information

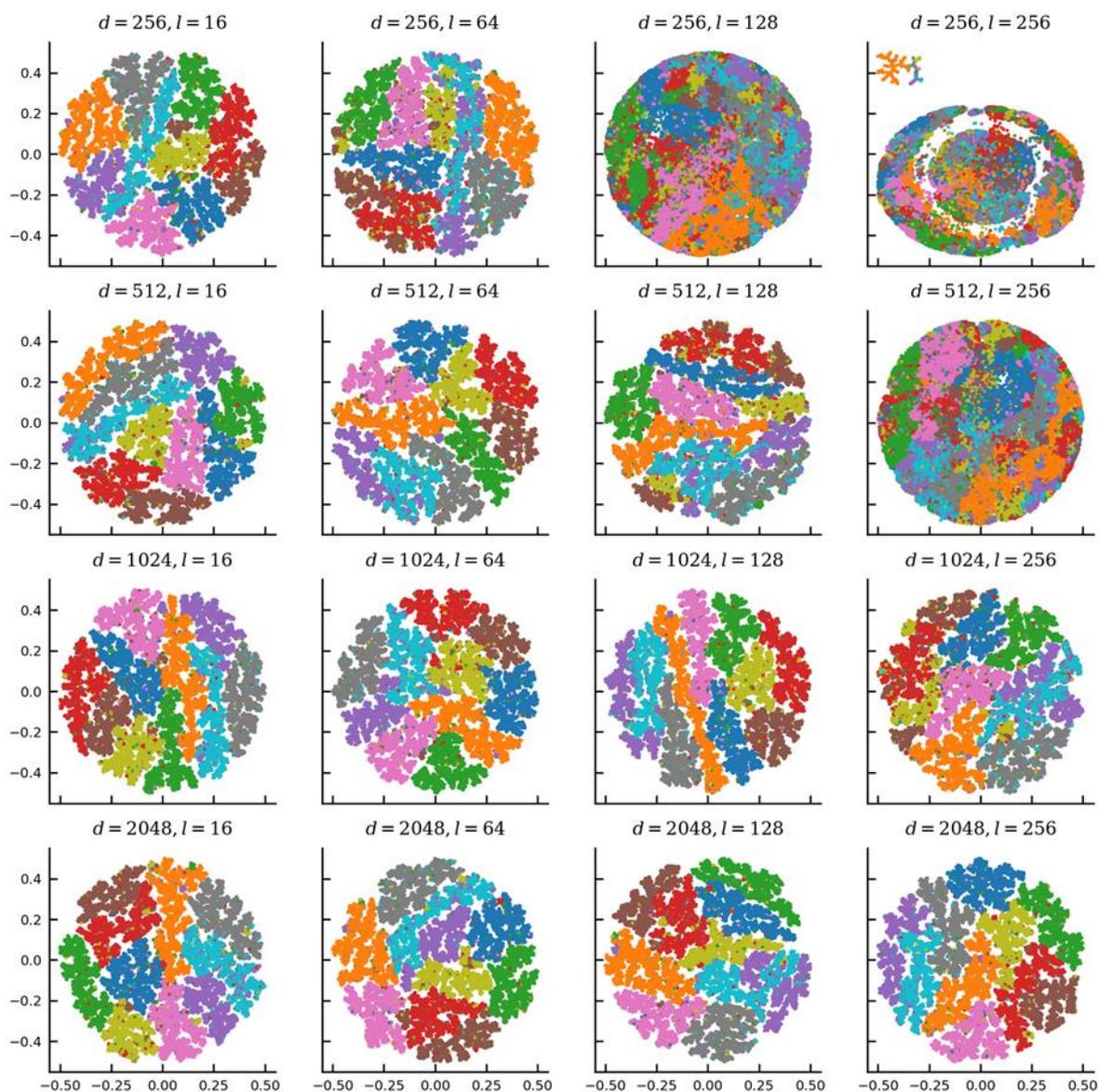


Fig. S1 Influence of LSH Forest parameters d and l on visualization of MNIST. While phase I of the algorithm mainly influences the preservation of locality (Fig. S6), extreme values where $d \approx l$ lead to a deterioration of visualization quality.

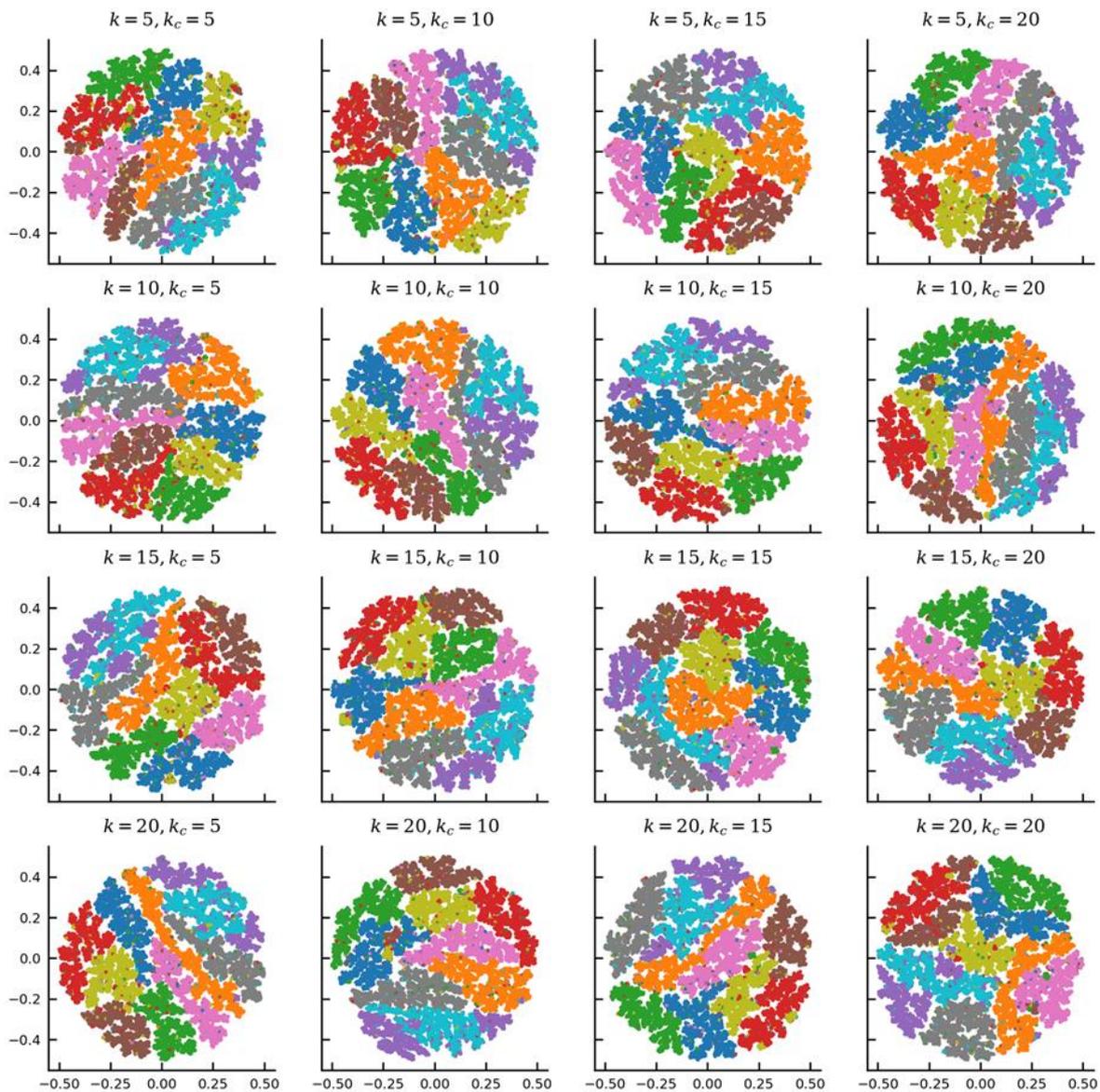


Fig. S2 Influence of LSH Forest parameters k and k_c on visualization of MNIST. Whereas parameter k directly influences the average degree of the k -nearest neighbor graph, k_c increases the quality of the returned k nearest neighbors. Both parameters only marginally influence the aesthetics and quality of the visualization.

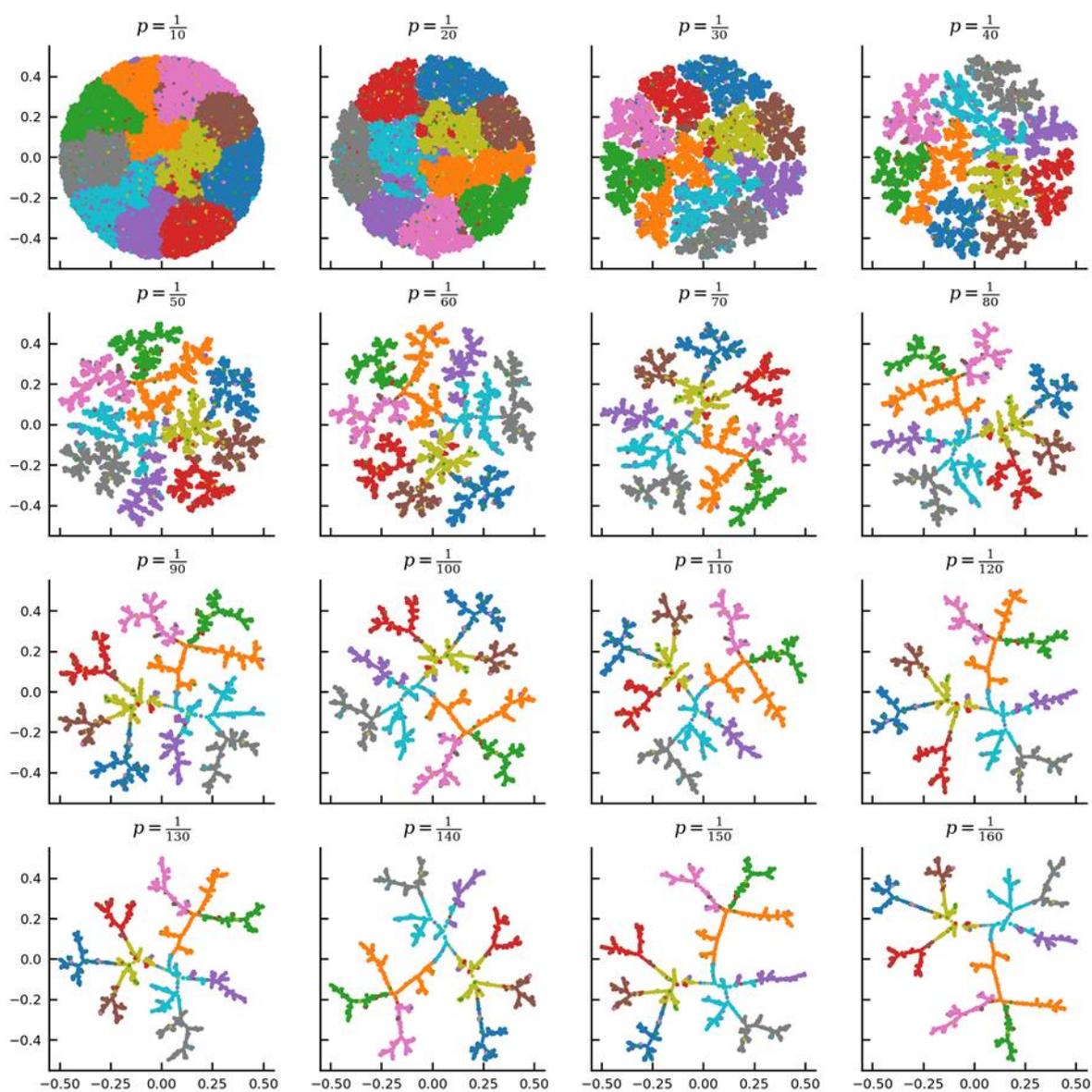


Fig. S3 Influence of parameter p on visualization of MNIST. The point size parameter p has major influence on the aesthetics of the visualization, as it controls the sparseness of the drawn tree. Decreasing the point size and thus the repulsive force between two points, allows the layout algorithm to draw points closer to their respective (sub) branches.

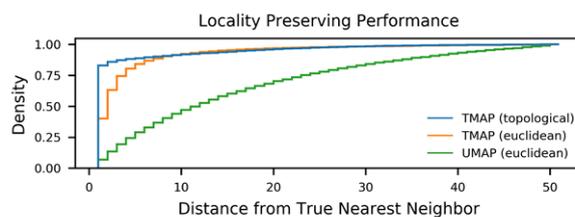


Fig. S4 Ranked distance from true nearest neighbor when visualizing the MNIST data set. Ranked distances from true nearest neighbor in original high dimensional space after projection based on topological and Euclidean distance for the MNIST data set. Whereas UMAP preserves less than 10% of true 1-nearest neighbors, TMAP preserves more than 80% based on topological and more than 35% based on Euclidean distance.

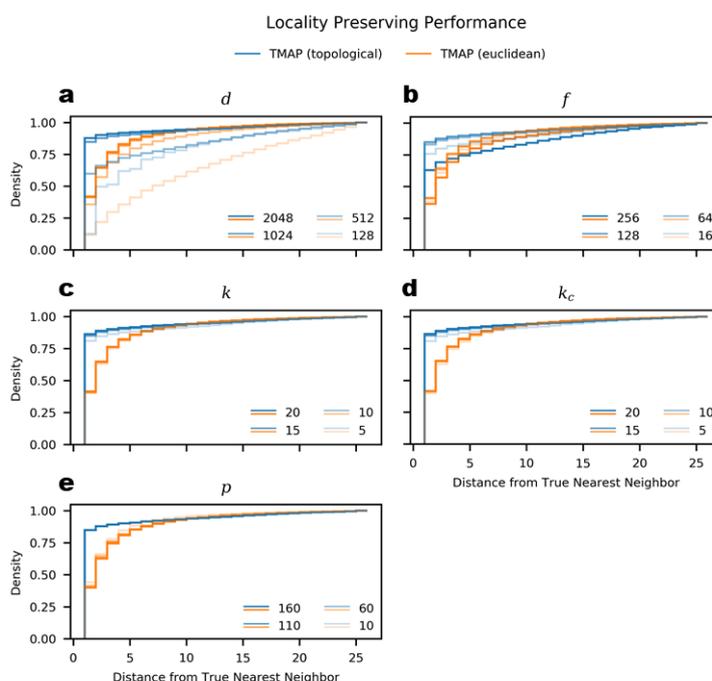


Fig. S5 Influence of TMAP parameters on locality preserving performance. Ranked distances from true nearest neighbor in original high dimensional space after projection based on topological and Euclidean distance for the MNIST data set. While, parameters d and l (a, b) have a major influence on both, the topological and Euclidean measure of locality preserving performance, parameters k and k_c have only marginal influence (c, d). The point size p does not influence topological distances; however, it has a minor effect on the Euclidean distance-based metric, as higher values increase the sparsity of the drawn tree.

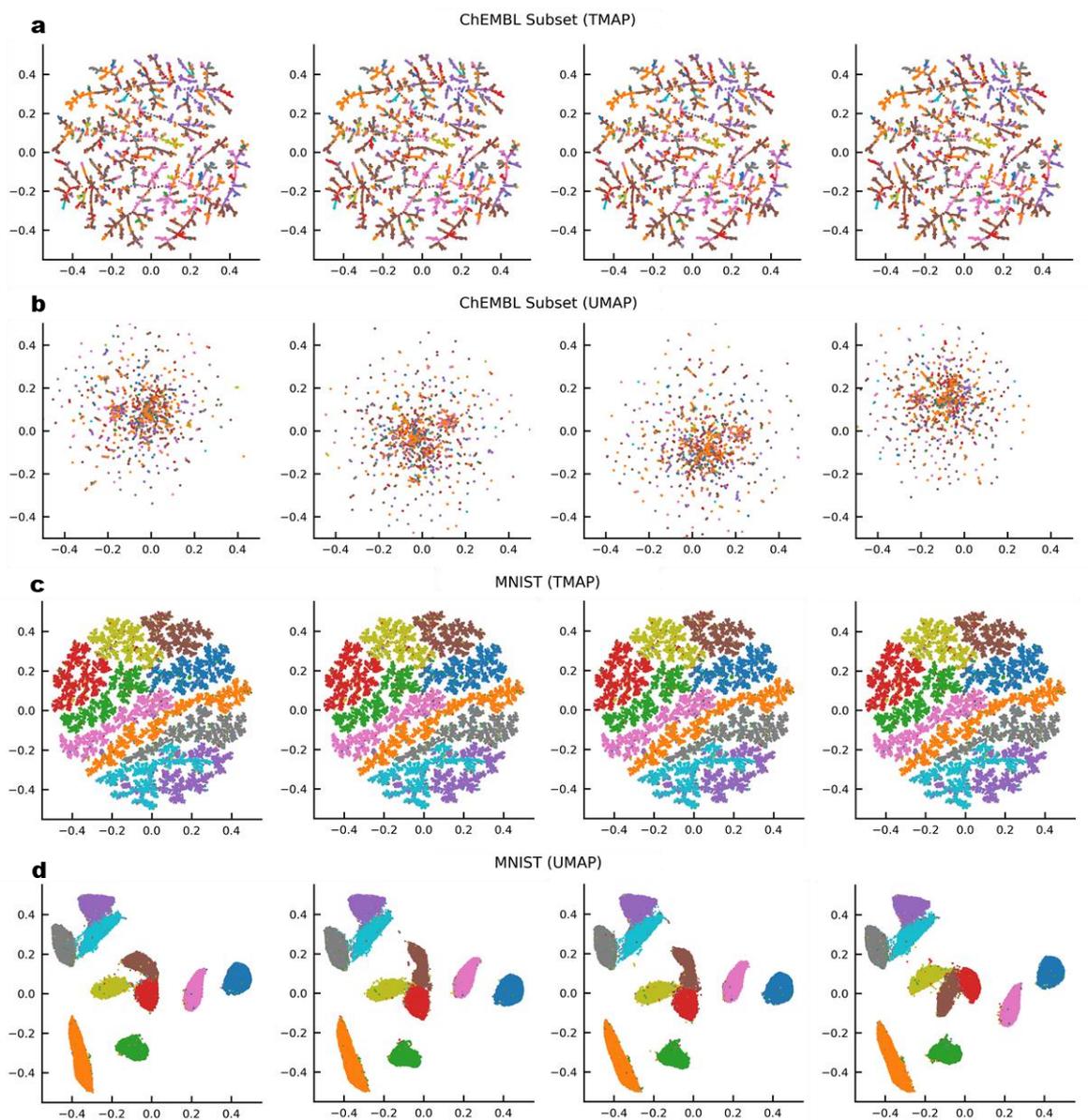


Fig. S6 Stability of TMAP. Algorithms TMAP (**a**, **c**) and UMAP (**b**, **d**) have been repeatedly ($n = 4$) run on the same data sets with the same parameters. Whereas the output of TMAP is perceived as identical in all instances, the results yielded by UMAP show considerable differences between each run.