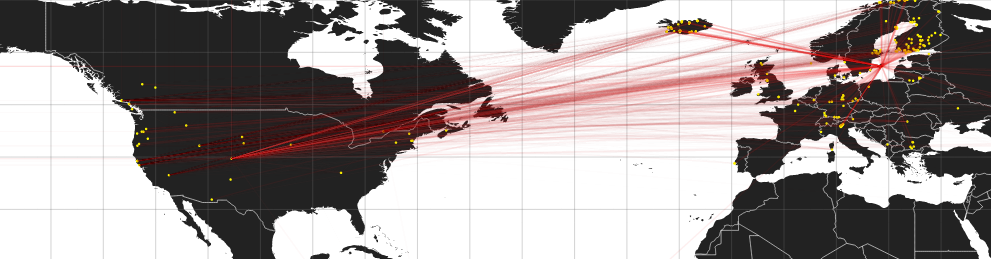
For this assignment, I have focused both on the visualization techniques and on implementing the data extraction feature from our database. Given the big analyzed database of strawberries that have been pinpointed to their extraction origin on the world map, cluttering became a serious problem even when analyzing smaller datasets (n~30 samples), because the number of edges increases exponentially with the node number.

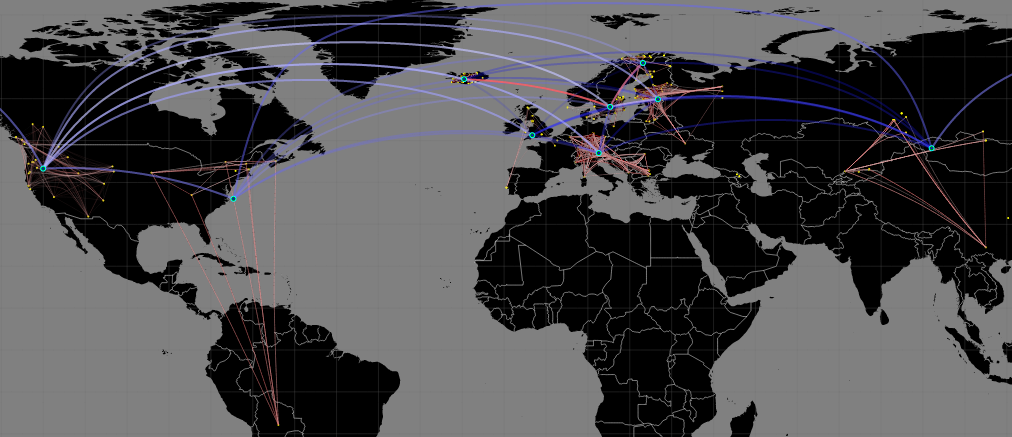
For reducing the cluttering with bigger data sets, I have used a literature algorithm for joining edges: Force Edge Bundling*[reference].* This algorithm reduces the area covered by edges and can provide a clearer image of the distance matrix mapped. While the image is not overloaded (Figure 1), regions with high sample density distract the viewer for other potentially important samples that are stranded in other less populated regions.



**Figure 1.** We can easily see that remote points from USA have strong connections to Europe’s samples, but distinguishing individual pairs is difficult.

For solving this issue, I have made a hybrid representation of a hierarchical visualization where points are clustered. Soft edges now connect points in the same cluster and, alongside with the force bundling algorithm, allows the viewer to focus on the key points.

For clustering, I have used an algorithm used in neural network classification. Therefore, the clusters are automatically created by the *K-Means Clustering* algorithm, implemented by *[reference]*. Good results have been obtained if the centroids are not predefined and the algorithm groups points without user input. After this part is done, I have implemented an algorithm for adapting the distance matrix to the centroids that represent the common characteristic of their underlying cluster. Last, the centroids are connected by bold arcs for a clearer representation.



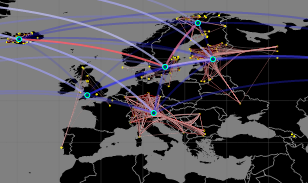


Figure 2. a) General visualization with interconnections in clusters and between centroids, where red shows strong connections between nodes. b) region of interest showing strong connections between clusters formed in Norway, Finland and Iceland (also confirmed from the research database