Reference: (Project from NUS CS5228 Knowledge Discovery and Data Mining)

The algorithm of Agglomerative Hierarchical Clustering is summarized as follows:

Input: Data points $X = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$; $x_i \in \mathbf{R}$ *and* $y_i \in \mathbf{R}$ are the coordinates.

Output: Clustering history: a list of pairs of the cluster ID, $H = \{(i_1, j_1), (i_2, j_2), \dots\}$, that indicates which pair of clusters are merged first; for example, $\{(1,3), (2,4), \dots\}$ indicates that (C_1, C_3) are merged first, then (C_2, C_4) are merged, ...

Steps:

```
1.    $C_i \leftarrow \{(x_i, y_i)\}, \text{for } \mathrm{i} \in \{1, \dots, N\}, \#\ current\ clusters \leftarrow N$
2.    Compute $ProximityMatrix[\mathrm{i}, \mathrm{j}], \text{for } \mathrm{i}, \mathrm{j} \in \{1, \dots, N\}$
3.    $ClusterIndexSet = \{1, \dots, N\}$ ; $H = []$
4.    Repeat:
          Find $(p, q)$ with $\underset{i,j \in \mathrm{ClusterIndexSet}}{\operatorname{argmin}} ProximityMatrix[\mathrm{i}, \mathrm{j}]$
          Merge $(C_p, C_q)$ together as $C_{N+1}$
          Update $ProximityMatrix[i, N+1], \text{for each } \mathrm{i} \in \mathrm{ClusterIndexSet}$
          Append $(p, q)$ into $H$
          $N \leftarrow N + 1$
          Remove $p$ from $ClusterIndexSet$
          Remove $q$ from $ClusterIndexSet$
          Insert $N$ into $ClusterIndexSet$
      Until $sizeof(ClusterIndexSet) = 1$
5.    Return $H$
```