Problem 2: Similarities between underground metro networks

We aim to investigate the similarities among various underground metro networks by analyzing their station connectivity and network structure. One of your collaborators has devised a metric distance measure for networks based on metro lines, station connections, and accessibility. The file `metro.txt` contains the matrix of the pairwise distances[1] between the underground metro networks of 90 different cities. In order to explore the similarities between these networks, we will employ a cluster analysis approach.

a) Which clustering methods discussed in class are suitable for this case?

b) Perform hierarchical clustering of the metro networks using the different linkages you have at your disposal. What is the number of clusters that should be discovered according to you? Which linkage(s) would you exclude?

c) Based on the cophenetic correlation coefficient, identify the best linkage.

d) Now, let us explore the DBSCAN method.

   Using a `minPts` value of 5 and choosing a consistent value for `eps` (with an accuracy of .05), run the DBSCAN algorithm. Is the result satisfactory?

e) Is there a way to visualize the underground metro networks in a two-dimensional plot? Report the plot, showing through it the results of the best clustering procedure identified. Assess whether the plot tends to underestimate or overestimate the true distances.

   Upload your results here: https://forms.office.com/e/q6m5j7peBu

---

[1]The matrix can be read as a classic table and then converted to an R `dist()` object with the command `as.dist()`