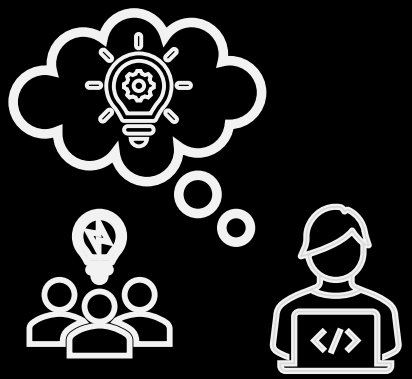
Lab session #6:
Clustering solutions comparison

Giulia Cisotto

Department of Informatics, Systems and Communication

University of Milan-Bicocca

giulia.cisotto@unimib.it



MOTIVATION

This sixth lab session aims to **compare and validate two different solutions of clustering (k-means, hierarchical)** applied to an unknown matrix of data (with low dimensionality and continuous attributes). This lab session refers to Prof. Stella's lectures no.5, 6 and 8 "Cluster Analysis: k-means", "Cluster Analysis: hierarchical clustering", and "Cluster Analysis: Clustering validation".

In this lab, you are going to (re-)use already developed code: *check solutions of Lab03, Lab04, Lab05*. Moreover, some new measures/strategies for validating clustering solutions will be used.

Today, **both hierarchical and k-means will be applied, with their own validation strategies**. Finally, comparison will be operated to finally guess the clusters, and cross-check with the true labels *(given at the end of the session)*.

For kmeans, two different strategies will be implemented:
1. running hierarchical clustering, find a candidate number of clusters, then run kmeans with that given number,
2. use the *elbow method* and find the best candidate number of clusters only using kmeans.

Motivation

Here is the link to **the Python code @Colab for today**

The **data** to work on will be **available on Moodle** at the beginning to the lab session.

The **true labels** to cross-check your clustering solutions will be **available on Moodle** at the end of the lab session.