

Measuring Clustering Quality

- ❑ **Clustering Evaluation:** Evaluating the goodness of clustering results
 - ❑ No commonly recognized best suitable measure in practice
- ❑ **Three categorization of measures:** External, internal, and relative
 - ❑ **External:** Supervised, employ criteria not inherent to the dataset
 - ❑ Compare a clustering against prior or expert-specified knowledge (i.e., the ground truth) using certain clustering quality measure
 - ❑ **Internal:** Unsupervised, criteria derived from data itself
 - ❑ Evaluate the goodness of a clustering by considering how well the clusters are separated and how compact the clusters are, e.g., silhouette coefficient
 - ❑ **Relative:** Directly compare different clusterings, usually those obtained via different parameter settings for the same algorithm