TD #8: Clustering

Large-scale Mathematical Programming

Leo Liberti, CNRS LIX Ecole Polytechnique
`liberti@lix.polytechnique.fr`

INF580

Summary

- Implement and test modularity clustering
- Implement and test cMINLP and MILP reformulations of MSSC
- Test random projections on cMINLP/MILP MSSC and compare with original formulations
- Verify distance resolution limit

Distance resolution limit

Use Python

- ▶ Test for dimensions $K \in \{5, 10, 50, 100, 1000\}$
- ▶ Generate random point sets X in K dimensions
- ▶ Evaluate distance matrix (using `scipy.spatial.distance.pdist`, see TD6)
- ▶ For each $x \in X$ compute min, max, avg, stdev over $\{d(x, y) \mid y \in X \smallsetminus \{x\}\}$
- ▶ Read a set of images into vectors (see TD6) and verify whether the distance resolution limit holds for these

MSSC formulations

Use AMPL

- ▶ Implement and test MP formulations for MSSC
    - ▶ convex MINLP (e.g. GUROBI, KNITRO, BonMin)
    - ▶ MILP (e.g. GUROBI, CPLEX)
    - ▶ Use a time limit!
- ▶ Both with and without random projections
- ▶ Compare all clusterings with mutual information
    see sklearn