MADR 2025 - Project tasks 6

Spectral clustering

Consider the following versions of the spectral clustering algorithm:

- Algorithm 1: **unnormalized spectral clustering**

input: a similarity graph $G = (V, E, W)$, number of clusters k

1. compute $L = D - W$ and its first k eigenvectors $u_1, \dots, u_k$,
2. let $y_1, \dots, y_n$ be the rows of the matrix containing $u_1, \dots, u_k$ as its columns,
3. cluster the points $y_1, \dots, y_n \in \mathbb{R}^k$ *to* $C_1, \dots, C_k$ by the k-means algorithm

output: clusters $A_1, \dots, A_k$: $A_i = \{j \in V : y_j \in C_i\}$.

- Algorithm 2: **normalized spectral clustering**

1. compute $\mathcal{L} = D^{-1/2} L D^{-1/2}$ and its first k eigenvectors $w_1, \dots, w_k$,
2. let $\phi_1, \dots, \phi_n$ be the rows of the matrix containing $D^{-1/2} w_1, \dots, D^{-1/2} w_k$ as its columns,
3. cluster the points $\phi_1, \dots, \phi_n \in \mathbb{R}^k$ *to* $C_1, \dots, C_k$ by the k-means algorithm,

output: clusters $A_1, \dots, A_k$: $A_i = \{j \in V : y_j \in C_i\}$.

- Algorithm 3: **normalized spectral clustering for $k = 2$ clusters, without k-means**

input: a similarity graph $G = (V, E, W)$, number of clusters k

1. compute $\mathcal{L}$ and its second eigenvector w_2,
2. set $y := D^{-1/2} w_2$ and order its coordinates,
3. find A_{opt} minimizing the Cheeger's cut among $A_1, \dots, A_k$ (as defined during the lectures)

output: clusters A_{opt}, A_{opt}^C.

- Implement the algorithms yourself.
- Use your implementations for several different data sets of your choice. Try different numbers of clusters to see the effect that this has.
- Compare the performances of the normalized and unnormalized algorithms on the same data sets.
- Can you provide an example in which Algorithm 3 (the one with some theoretical guarantees) is superior to Algorithm 2 (with two clusters)?