HW4: Graph Spectral Analysis (K-Eigenvector)

Authors: Andrei Capastru, Alex Porciani | Group 11

The task is K-Eigenvector algorithm from the paper "On Spectral Clustering: Analysis and an algorithm" by Andrew Y. Ng, Michael I. Jordan, Yair Weiss, in order to find "neighborhoods" - groups of nodes or clusters.

1.1 How to run the code

0. Install python3 and install Jupyter Notebook
1. Run in unix terminal:

```
cd ~
git clone https://github.com/andreicap/data-mining.git
cd data-mining/HW4
jupyter notebook
```

2. Open the browser on `http://localhost:8888` and run the `graph-spectra.ipynb` file

1.2 Calculating the Graph Matrix Laplacial

In this part of the code we calculate the **Matrix Laplacial** - which we will use later on to cluster the vertices.
Steps taken to calculate the Laplacial:

1. Create the `adjacency matrix` A - 1s in position `{i,j}` if there is a connection and 0s otherwise.
2. Create the diagonal matrix D, with vertical sums of the columns of `A` on the main diagonal.
3. Compute the Laplacian Matrix $L = D^{-\frac{1}{2}} \cdot A \cdot D^{-\frac{1}{2}}$.

1.3 Clustering

Steps taken to cluster the graph vertices:

1. Extract the eigenvalues w and eigenvectors v of L.
2. Extract highest values with difference 0 from w, and count them into k .
3. Extract first k columns of eigenvectors from v and renormalize it.
4. Use kmeans to with the expected number of clusters k to group the vertices into their respective cluster.

Output example: