Data Mining 01.12.2017 – Piotr Mrowczynski – Assignment 4

The objective of this assignment was to do spectral graph analysis and clustered points using algorithm as described in the paper “On Spectral Clustering: Analysis and an algorithm”. Using implementation of the K-eigenvector algorithm, we analyse two sample graphs.

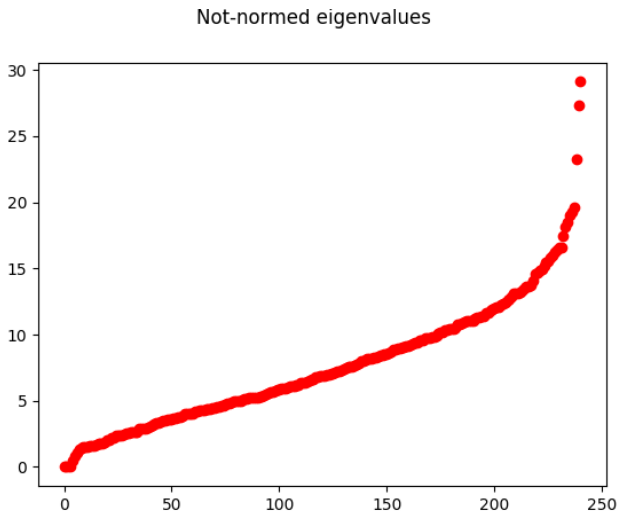
Project structure

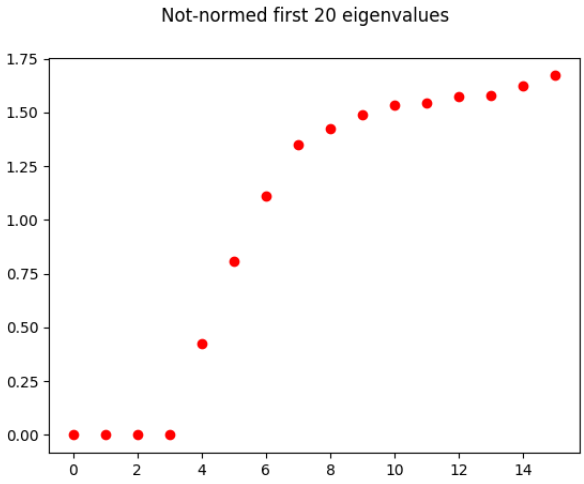
```
graph-spectra
├── spectra.py          : Algorithm making analysis of spectrum for specific graph
└── GraphSpecta.odt
```

Running

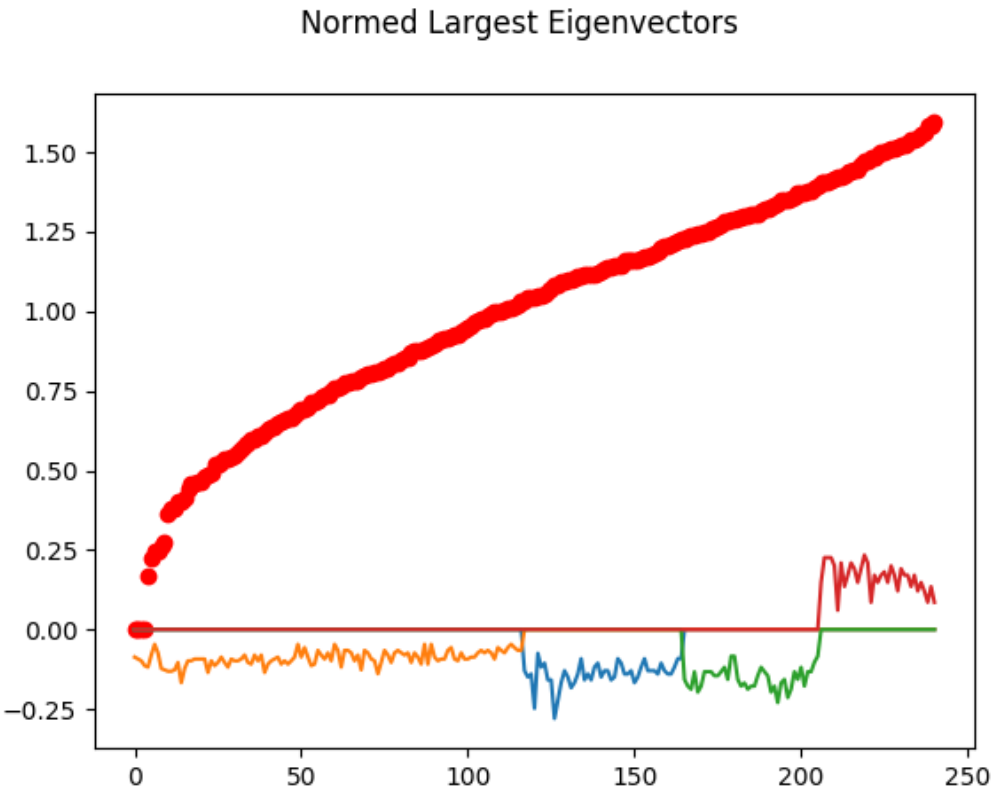
python graph-spectra/spectra.py graph-spectra/example1.dat 4
python graph-spectra/spectra.py graph-spectra/example2.dat 1

Results graph 1 example1.dat:

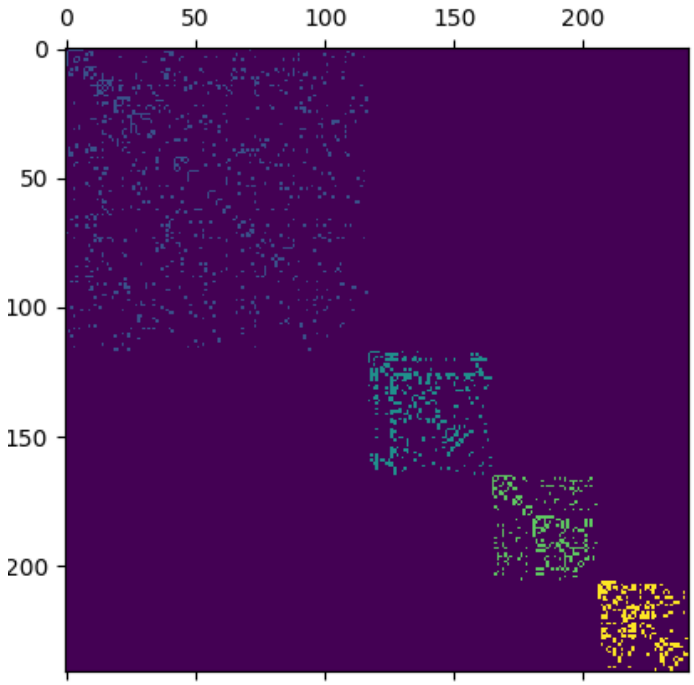




From the spectrum of not-normed eigenvalues of the graph, we can see that this graph will have around 4 cluster. We can deduce it from the eigengap at 4^{th} eigenvector.

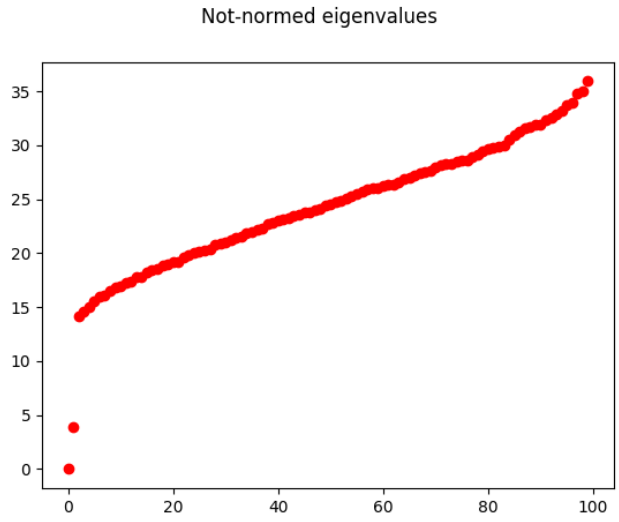


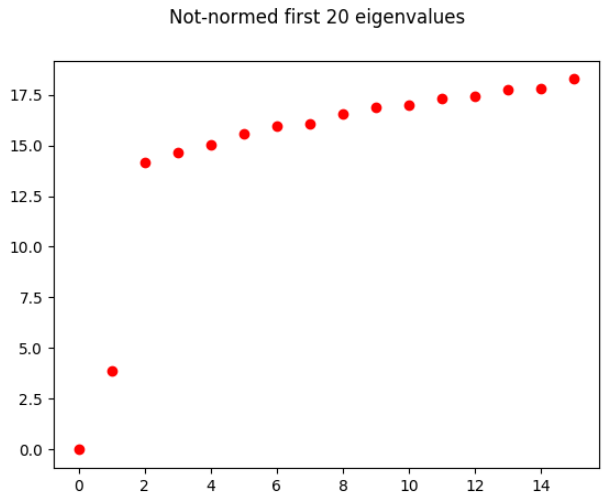
From the spectrum of 4 largest eigenvectors, and normed eigenvectors spectrum, we can deduce that this graph is disconnected



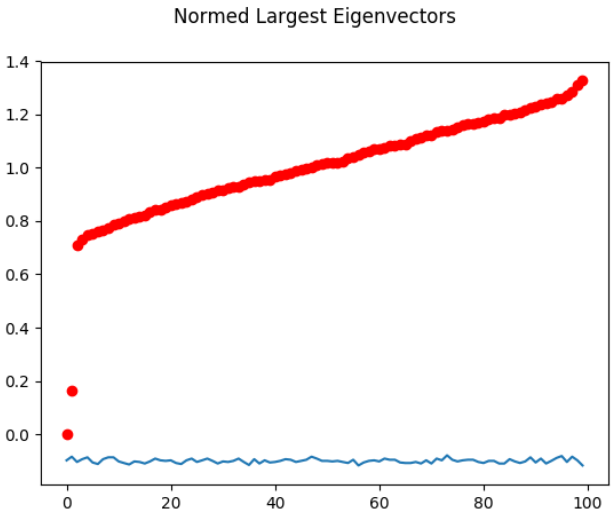
Algorithm, based on spectrum of 4 largest eigenvectors, using k-means of eigenvector point's in 4 dimensions (a,b,c,d), k=4, correctly identified 4 clusters

Results graph 2 example2.dat:





From the not-normed eigenvalues spectrum, we can see that this graph most probably will have 1 or 2 clusters.



Additionaly, we can see that its second eigenvalue is not much different from 2nd eigenvalue, which suggests very big connectivity and large cluster.