

# Data Mining: Frequent Itemsets and Association rules

Braulio Grana Gutiérrez      Adrián Ramírez del Río

December 18, 2016

## Description

For this assignment we had to implement the Spectral Clustering algorithm [1]. To do so, we used MATLAB and the *Statistics and Machine Learning Toolbox*.

We created a function that receives a list of edges  $E$  and the number  $k$  as parameters, and implemented the algorithm with the following steps:

1. We use the adjacency matrix  $A$  as affinity matrix.
2. Build the degree matrix  $D$ , which is a diagonal matrix whose  $(i,i)$ -elements are the sum of  $A$ 's  $i$ -th row.
3. Build the normalized Laplacian matrix  $L$  using the following formula:

$$L = D^{1/2} \times A \times D^{1/2}$$

4. Build matrix  $X$  by finding the  $k$  largest eigenvectors of  $L$  and stacking them in columns. For this we used the MATLAB method *eigs*.
5. Renormalize each row of  $X$  to have unit length (Store the result in  $Y$ )
6. Use K-means from the *Statistics and Machine Learning Toolbox* to find the  $k$  clusters in  $Y$ .
7. Assign each original point to its cluster and return the result  $C$ .

Our function returns a vector  $C$  with tuples containing a point and its assigned cluster.

## Instructions

To use our implementation:

1. Open MATLAB and change the current folder to the project root.
2. Then, in the MATLAB console, execute the following commands:

```
E = readcsv( 'Path_to_dataset' );  
C = SpectralClustering(E, k);
```

## References

## References

- [1] Andrew Ng, Michael I. Jordan and Yar Weiss, *On Spectral Clustering: Analysis and algorithm*, 2001.
- [2] Ulrike von Luxburg, *A tutorial on Spectral Clustering*, 2007.