Introduction

Clustering is one of the most used tools for exploratory data analysis. It is widely applicable in numerous scientific fields: from computer science to psychology. Due to popularity of the technique, extensive studies have been made and multiple clustering algorithms were developed. One of them, spectral clustering, is of our particular interest.

Popular clustering techniques such as K-means or mixture-model learning work well in cases where some assumptions regarding data are imposed. For instance, K-means works well only if data points are located in such a way that they form a convex set. To apply mixture-model learning one should assume that the data has a particular distribution. In real-world problems data is often very heterogeneous, so these methods often demonstrate poor results. Spectral clustering on the other hand tries to represent data in an easy-to-group way so that traditional algorithms will trivially find the correct clusters.

A referenced algorithm mostly relies on spectral analysis of affinity matrix of the data. There are several variations of this algorithm (references to Ng, Malik…) which have small differences mainly in computing Laplacians. Also, many improvements were proposed (self-tuning paper, …) and are up to discussion.

(MAYBE) A technique for automatic scaling parameter selection sigma will be proposed and discussed.

(ADD DESCRIPTIONS OF FUTURE SECTIONS)