

SpectraLib_A (DRAFT as of June 06, 2007)

Marina Meila
mmp@stat.washington.edu

Tatiana Maravina
maravina@stat.washington.edu

June 06, 2007

1. Introduction

This toolbox is an extension to the SpectraLib package written by Deepak Verma. It mainly consists of two parts:

- Code to generate or read the data (in the data subdirectory)
 - Symmetric and asymmetric "similarity matrices" that normalize to block stochastic P's (the algorithms used are described in [2])
 - Different patterns in 2D (clusters presented by circles, stripes, mixture of Gaussians)
 - Reading and preparing the data for asymmetric clustering (based on WEBKB example, see [1])
- Code to run the WCut algorithm for asymmetrical clustering presented in [1] (in the 'asym code' subdirectory)

The main directory contains `install_SpectraLib_A.m`, the code which adds all the library directories to the Matlab session. The two main subdirectories, as described above, are `data` and `'asym code'`. The helper subdirectory contains helper functions. There is usually an example in the comments to each function of how to use it.

2. Using the Toolbox

2.1 Starting the package

To install the SpectraLib_A set ASYMSPECTRAL_HOME to a string containing the name of the directory in which the package was saved and make sure that install_SpectraLib_A.m is in a directory which Matlab can access. This package uses functions from the Spectral toolbox (by Deepak Verma) so it must also be installed. To obtain it go to: www.ms.washington.edu/~spectral/. If it is already installed in the current Matlab session nothing needs to be done. If it has not been installed then SPECTRAL_HOME should be set to a string containing the name of the directory where it has been saved. With this information both libraries will be installed by running the small script intall_SpectraLib_A.m.

2.2 The structure

ASYMSPECTRAL_HOME/helper

GraphComponent.m

Sexp_from_points.m

VI.m

cluster_normalized_kmeans.m

connected_components.m

getlargestcomponent.m

telescope.m

untelescope.m

ASYMSPECTRAL_HOME/asym code

cluster_wcut – function that implements the BestWCut algorithm as described in Marina Meila and William Pentney 'Clustering by weighted cuts in directed graphs'

wcut_main.m – a script that runs WCut with many different T (weights) and selects the result by distortion

ASYMSPECTRAL_HOME/data/read

`webkb_readdata.m` – a script to read the asymmetric Web Graph Data stored in `webkbdata.mat`

ASYMSPECTRAL_HOME/data/generate

`genBlockACitationMatrix.m`

`genBlockBandMatrix.m`

`genBlockDiagMatrix.m`

`genBlockEqRowsMatrix.m`

`genBlockMatrix.m`

`genBlockSumMatrix.m`

`genBlockSumSymMatrix.m`

`genLinkBlock.m`

`genPerfectH.m`

`gen_2D_circles.m`

`gen_2D_mixgauss.m`

`gen_stripes.m`

`gen_stripes_lattice.m`

`gen_sym_hypercube.m`

3. References

[1] Marina Meila and William Pentney “Clustering by weighted cuts in directed graphs”, SIAM Conference on Data Mining (SDM) 2007

[2] Marina Meila and Tatiana Maravina “Methods for generating matrices with PCE”, [in progress](#)