## Utilization of constrained spectral clustering for clustering of graph nodes containing record data\*

## László Gonda<sup>a</sup>

<sup>*a*</sup> University of Debrecen, Faculty of Informatics gonda.laszlo@inf.unideb.hu

## Abstract

Clustering is one of the most common and most widely used methods of data mining. Many clustering algorithms can be utilized for various purposes, but most of these methods can only deal with one data type at a time. There are few methods in existence that can deal with data of different origins and/or different types simultaneously, but most of these are iterative [1], which means that they consume large amounts of computation time, and therefore refreshing the clustering can be a serious issue as well.

However, by using the method of constrained spectral clustering, which was originally established to define graph nodes that surely fall into the same cluster or surely not, we can incorporate the influence of the similarities of regular record data associated to the nodes without using complex iterative methods. Furthermore, using this system, we can also introduce different "constraint matrices" that can represent different aspects of the objects, or different subsets of their attributes grouped together based on importance, etc. and incorporate them in the forming of the final clustering without the need for an even lengthier iteration, or serious alterations to the original system. The original constrained spectral clustering method in [2] can also be weighted to determine whether the constraints or the graph edges will be of higher importance in the resulting clustering, thus we can say that the method is customizable in many aspects.

*Keywords:* data mining, clustering, constrained spectral clustering, graph data, record data, co-clustering

MSC: 68U99

## References

- [1] WANG, JIDONG, et al.: ReCoM: Reinforcement Clustering of Multi-Type Interrelated Data Objects, *Hong Kong University of Science and Technology* (2003)
- [2] WANG, XIANG: Spectral Clustering for Complex Settings, *Dissertation, University of California Davis* (2013)

<sup>&</sup>lt;sup>\*</sup> This research is supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund. supported by the European Union, co-financed by the European Social Fund.