

BDA group master project

Project title: In-silico testing of data fusion methods
Duration: 6 months
Number of ECs:
Start: negotiable
Supervisor: Age Smilde and Evrim Acar
Contact: a.k.smilde@uva.nl

Project description

In the life-sciences, the number of data sets measured on the same system is growing. These can be, e.g., RNAseq data, proteomics and metabolomics data. This then generates multiple sets of data that are connected and should be analyzed simultaneously. The class of methods dealing with such analyses is called data fusion methods. Sometimes these measurements are also performed at several time points. There is a growing number of methods available in data fusion but the methods that also take into account the time axis is limited.

Because of the availability of the time axis, some of the data sets are now three-way arrays (e.g., samples, biochemical compounds, time), and data fusion methods that can jointly analyze such multi-way arrays with matrices (or multiple multi-way arrays) are called for to deal with these data sets. We formulate data fusion as coupled matrix and tensor factorizations (CMTF), and also consider extensions of CMTF to incorporate the time-evolving nature of the underlying patterns in dynamic data sets. It is not easy to test the adequacy of these methods in real-life situations and that is where in-silico models can help. Specifically, the single-cell in-silico model of Karr¹ can be used to generate time resolved RNA, proteome and metabolome data. Since the ground-truth is known, the methods can be tested for their relevance.

This project is a collaboration between the BDA group of SILS (Age Smilde) and Simula Research Laboratory in Norway (Evrin Acar). Different types of data sets will be generated using the Karr model and these data sets will be analyzed using coupled matrix and tensor factorizations (using the available methods in the CMTF Toolbox² as well as their extensions to time-evolving data). Then conclusions can be drawn as to the applicability of the methods.

¹J.R. Karr et al; A Whole-Cell Computational Model Predicts Phenotype from Genotype, Cell 150, 389–401,2012

²Available at http://www.models.life.ku.dk/joda/CMTF_Toolbox