

Two-Tier Mapper: a user-independent clustering method for global gene expression based on topology

Rachel Jeitziner

Swiss Institute for Experimental Cancer Research,
School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne,
CH-1015 Lausanne, Switzerland

`rachel.jeitziner@epfl.ch`

I will present a new method to analyze biological data, based on topological data analysis. I will introduce the mathematics underlying this new tool, illustrate its utility through a real data example of its application, and outline theoretical aspects of its stability.

This method provides a first approximation to the variability in a dataset, describing divergences from sample to sample. It clusters samples both globally and locally according to a function of interest and provides therefore more insights into the data. It comprises a visualization tool that distinguishes the various clusters, giving an easy-to-grasp presentation of the variation between samples in the dataset as a colored graph.

The method, which is based on the well-known Mapper algorithm, can be applied reliably to both small and large datasets, which is a clear advantage in comparison with standard statistical tools, which perform reliably only on datasets of at least a certain minimal size. All parameters are determined either in a data-driven manner or by choosing reliable, user-independent default parameters.

The method is developed as an open source package in R deposited at the Bioconductor.

Joint work with:

Mathieu Carrière, Jacques Rougemont, Steve Oudot, Kathryn Hess, and Cathrin Brisken