

# Visualizing metadata change in networks and / or clusters

Tanguy LALLEMAND<sup>1</sup>, Sylvain GAILLARD<sup>1</sup>, Sandra PELLETIER<sup>1</sup>, Claudine LANDÈS<sup>1</sup>, Sébastien AUBOURG<sup>1</sup> and Julie BOURBEILLON<sup>1</sup>

<sup>1</sup> IRHS, Agrocampus-Ouest, INRA, Université d'Angers, SFR 4207 QuaSaV, 49071, Beaucouzé, France

Corresponding Author: [julie.bourbeillon@agrocampus-ouest.fr](mailto:julie.bourbeillon@agrocampus-ouest.fr)

## 1 Introduction

Among current trends in various research fields, in particular in biology, is the increase in the scale at which studies are performed. This results from the wider spread of high-throughput experimental techniques such as transcriptomics or proteomics and the increase of the volume of publicly available datasets. For instance, the biology teams from the IRHS (Institut de Recherche en Horticulture et Semences) in Angers have been accumulating datasets of different natures (transcriptomic, biochemistry, physical measures, sensory analysis, etc.) regarding perennial, annual and biannual plants. However experiments are performed independently and resulting data are cross-analysed manually and a-posteriori by scientists [1]. Therefore the demand by biologists to integrate heterogeneous and large datasets from "omics" and phenotyping activities is rapidly increasing [2]. The classical approaches imply collating various data sets into a large data matrix and mine the matrix through methods such as building networks to represent relationships between items or clustering items into closely related groups. These are in turn visualized as graphs or heatmaps for instance.

## 2 Issue at hand

The first problem is the interpretation of the visualizations in biological terms. A standardized description of each dataset in the integrated matrix has to be available. These metadata are more and more well-defined through standard formats such as MIAME [3] and filled in with concepts from reference ontologies. However those have to be presented to the user in an interpretive way, which can be challenging when dealing with annotations extracted from large knowledge representations such as the GO [4]. Moreover building networks or clustering are generally iterative approaches. Kinetics are also regularly found among biological datasets. The presentation of the metadata is then not only a one shot operation but has also to take into account some kind of chronology through the data mining steps or the course of the biological process.

## 3 Contribution

In this context we are developing a web-based tool to present biologists with visualizations of ontology annotations associated with biological network graphs or cluster heatmaps through time including: (i) as a series of snapshots corresponding to successive steps and (ii) a representation of the difference between two steps. Our approach builds on methods such as GO enrichment analysis [5] and visualization of changes in networks [6].

## References

- [1] Mercedes Arguello Casteleiro et al. Deep learning meets ontologies: experiments to anchor the cardiovascular disease ontology in the biomedical literature. *Journal of Biomedical Semantics* (9):13. 2018.
- [2] James Hendler. Data Integration for Heterogenous Datasets. *Big Data*. 2(4):205-215. 2014.
- [3] Alvis Brazma et al. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nature Genetics* (29):365–371. 2001.
- [4] Michael Ashburner et al. Gene Ontology: tool for the unification of biology. *Nature Genetics* (25):25–29. 2000.
- [5] Huaiyu Mi et al. Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols* (8):1551–1566, 2013.
- [6] Martin Rosvall and al. Mapping change in large networks. *Plos One*. 5(1): e8694 2010.