

BioNetStat: A differential network analysis tool to biological data

Vinicius Jardim Carvalho¹, Suzana de Siqueira Santos², Andre Fujita², Marcos Silveira Buckeridge¹,

1 University of São Paulo

2 IME - USP

Abstract

The networks theory is an important way to model and understand the interactions diversity of biological systems, considering from cells organelles to the whole biosphere. The dynamic of systems structure, such as the changes in the interactions among the system elements, is an inherent trait of those systems. To represent each one of the many states assumed by a system we can use networks. In this sense, there is a wide range of tools proposed to compare those networks. However, none of them are able to compare structural characteristics among more than two networks. Considering that systems generally assume more than two states, we developed a statistical tool to compare more than two networks and highlight key players in the process studied. The main proposition of this study was to compare correlation networks using traits that are based on graph spectra (eigenvalues set of adjacency matrix), such as the spectral distribution. This measure is associated with other traits of networks, such as the number of walks, diameter, and cliques, and it is a better characterization of graphs than other classical measures of networks theory. In addition to spectral distribution, we also compare networks by spectral entropy, degree distribution, and nodes centralities. To verify the performance of tool we used a tumoral cells genes expressions data set. In the case studies, we used two data sets, the same gene expression data and a plant metabolites concentrations data. The method proposed, called BioNetStat, was implemented in R package with a user interface for people that do not programming. We verified that the method is efficient to distinguish more than two networks. However, the increase in networks number and the decrease in sample unities reduce the statistical power of methods. The method proposed brings a potential time economy, doing a single analysis to compare more than two networks rather than compare them by pairs. The method highlighted sets of variables with a central role in biological systems that were not highlighted in other studies where only gene expression or metabolic concentration were analyzed. In that way, we proposed another way to find traits (variables) that distinguish cancer types genes expression or organ plants metabolisms. Furthermore, the highlighted variables allow us to create hypotheses about its role in the process studied, bringing new finds about the systems operation mechanisms.

Funding: FAPESP, CAPES