ANOVA-like method for differential correlation of multiple networks analysis of biological data

Vinícius Jardim Carvalho 1,2, Suzana de Siqueira Santos 3, Adriana Grandis 4, Amanda Pereira de Souza 5, André Fujita 3, Marcos Silveira Buckeridge 2

1 *Bioinformatics Graduate Program, University of São Paulo, São Paulo, SP, Brazil,* ***2*** *Department of Botany, Institute of Biosciences, University of São Paulo, São Paulo, Brazil,* ***3*** *Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil,* ***4*** *Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA,* ***5*** *National Institute of Amazon Research, Manaus, Brazil*

Identify if metabolites or genes expression patterns range over an experiment in response to environmental conditions is a major task in Bioinformatics. Therefore statistical tests such as t-test and ANOVA, are used to identify which variables significantly range between two or more biological conditions. However, those tests do not take into account the information about the relationships among variables. To overcome that limitation, several multivariate methods were developed such as PCA, cluster analysis, multiple regression, and network analysis such as the CoGA software. Network analysis allows us to address the connectivity between studied variables. The CoGA software performs differential network analysis between two graphs (one for each biological condition) based on network topological characteristics, such as centralities, clustering coefficient and spectral distribution. Despite the important role of CoGA, plant physiology experiments often compare more than two biological states. In order to fill this gap, we aim to implement a generalization of the CoGA method for two or more graphs, which can be very useful when we compare many biological states.The ANOVA-like test for several graphs performs based on the mean of Kullback-Leibler divergences between their spectral distributions (distributions of the eigenvalues of the graph adjacency matrices) and the average distribution of all graphs. The pvalue of statistical test for the divergence is made through permutation of the labels. In order to evaluate if the spectral distribution test controls the false positive rate and to measure its statistical power we performed simulation experiments with biological data. Thus the proposed method can bring evidences of differences in the network structure among two or more biological conditions. The application of our method to two sets of data on plant physiology and biochemistry obtained from a C4 and a C3 plant under different experimental conditions revealed that the methodis reliable for robust statistical comparisons among networks eitherwithin 24h or along several weeks. We expect that this method will beuseful for plant (and animal) physiologists, providing means to analyzelarge data sets that integrate molecular, biochemical and physiologicaldata so that whole organisms could be scientifically compared under different environmental conditions.