statGraph: a statistical tool to analyze biological networks

Suzana de Siqueira Santos[1], Daniel Yasumasa Takahashi[2], Andre Fujita[1],

1 IME - USP
2 Department of Psychology and Neuroscience Institute, Princeton University, Princeton, United States of America

Abstract

Several biological systems can be modeled by graphs (networks), which represent the interactions/relationships among the elements of the system. Examples of systems represented by graphs include gene regulatory networks, protein-protein interaction networks, and brain functional networks. Understanding how components interact with each other and the changes that occur in the system under certain conditions is a major concern in several fields of biology. One challenge of biological system studies is dealing with their variability. For example, the coactivation between regions of the brain changes across time, and gene regulatory networks are not identical even among individuals that share the same phenotype. Therefore statistical methods that model the randomness/variability of the graph structure are essential to analyze biological networks. However the available computational tools to analyze graphs lack statistical methods (frequently they do not include any statistical test or include tests for only few specific features of the graph). Here we present an R package, namely statGraph, that includes several statistical methods to analyze graphs, such as a test to compare the structure among two or more sets of graphs, a test to verify whether two sequences of graphs are correlated, a model selection approach and a procedure for parameter estimation. All methods included in statGraph are based on the spectrum of the graph (set of eigenvalues of the adjacency matrix), which is associated with several properties of the graph, such as number of walks, cliques and diameter. The methods were validated through simulation experiments, which show that they behave as expected for graphs with sufficient number of vertices.

Funding: S.S.S. was partially supported by CAPES and FAPESP (2015/21162-4). A.F. was partially supported by FAPESP (2013/03447-6, 2015/01587-0, 2016/13422-9), CNPq (304876/2016-0), NAP eScience-PRP-USP.