

Pattern Recognition in genomic sequences: A case of study using complex networks

Isaque Katahira^{1,2}, Fabrício Martins Lopes¹

¹*Universidade Tecnológica Federal do Paraná – Campus Cornélio Procopio
Departamento de Computação*

Programa de Pós – Graduação em Bioinformática – PPGBIOINFO

²*Centro Paula Souza – Escola Técnica Estadual Professor Mário Antônio Verza*

The integration of computing with other knowledge fields is becoming increasingly important for the research development in various fields of science. Notably, Bioinformatics is an example of this fact, since it is inherently multidisciplinary by combining knowledge of biology, statistics and computer science among other relevant sciences. In this context, the simultaneous extraction of molecular data from thousands of genes is a major breakthrough in the area, but also a challenge, as it generates a huge volume of biological data to be analyzed. In particular, it is essential to develop new methods and techniques in order to understand the influence of gene expression in the functional state of an organism. In this context, a possible way that can contribute is to develop techniques able to reduce the amount of data without loss of information contained therein. In this sense, the feature extraction techniques can be applied, which can be used to represent the data through its most relevant properties without using the entire data set. Thus, it is proposed to develop an approach based on mapping of genome sequences in their complex networks representation. From these networks can be taken measures to specify them into feature vectors which may be used to summarize the data collected, in order to quantify the topological similarity between the generated networks. It is justified, therefore, the use of complex networks for its approach to the real networks with respect to nonlinearity, where structural arrangements of its nodes can be crucial for understanding the interactions, functions and its whole structure. The observation of the extracted characteristics via nodes, edges, and other arrangements may provide the identification of patterns contained in these networks. Therefore, it is expected that the network measures leads to distinguish different classes of genomic sequences, turns possible to verify mutations, help in constructing phylogenetic trees, and other analyzes. Thus, there is the potential to observe different patterns contained in biological structures and propose large-scale rating of applications in the context of systems biology, that is, considering the modeling of an organism as a whole.