

Gene functional similarity analysis based on the Gene Ontology

Evaluation of the functional similarity between genes or gene products is a central issue in bioinformatics. Computational tools are required in order to cope with the huge amount of data available. In addition to traditional tools, e.g. those for the analysis of nucleic or amino acid sequence similarity, functional similarity of genes or gene products can be evaluated also by analyzing their functional annotations available in several databanks. For this aim, functional annotations expressed by using controlled vocabulary terms, either unstructured or organized as an ontology, i.e. with defined semantic relations between them, are very useful. In particular, among them, the Gene Ontology represents the *de facto* “standard” reference for annotation-based functional similarity analyses.

The primary goal of the proposed thesis is to design and implement a software platform that enables the computation of functional similarity measures between genes or gene products on the basis of their annotations with the Gene Ontology and other bio-terminologies e bio-ontologies. Since in the literature examples of software libraries developed to this aim exist, a preliminary study of the available off-the-shelf solutions is required in order to evaluate the chance/opportunity of integrating such libraries within the software platform to be developed. Due to the large amount of data to be processed by the algorithms involved, the thesis will also have to study and optimize the access methods to the data sources available.

A further, goal of the proposed thesis is to complement the functional similarity scores, obtained with state-of-the-art similarity metrics, with innovative algorithms inspired by the literature on complex networks. In fact, by modelling the genes or gene products and their controlled annotation terms (e.g. their Gene Ontology annotation terms) as nodes of a network, this thesis will investigate the use of eigenvector centrality methods (closely related to Google Page Rank) to define new similarity scores based on the local connectivity of the modelled network. Result validation will be performed by evaluating the computational solutions obtained performing gene clustering or gene ranking based on the calculated functional similarity scores.

Contacts:

Marco Tagliasacchi

marco.tagliasacchi@polimi.it

Tel: 02 2399 7680

Marco Masseroli

masseroli@elet.polimi.it

Tel: 02 2399 3553