

Co-expression network games

Bonassi S⁽¹⁾, Fragnelli V⁽²⁾, Moretti S⁽³⁾, Patrone F⁽⁴⁾

⁽¹⁾ Unit of Molecular Epidemiology, National Cancer Research Institute, Genoa

⁽²⁾ Department of Advanced Sciences and Technologies,
University of Eastern Piedmont, Italy

⁽³⁾ Unit of Molecular Epidemiology, National Cancer Research Institute, Genoa

⁽⁴⁾ DIPTeM, University of Genoa, Genoa

Motivation

Gene co-expression networks and other biological networks (e.g. representing protein-protein interactions) are increasingly used to explore the system-level functionality of genes and proteins. For instance, the interpretation of gene interaction in co-expression networks, generates the need for a meaningful ranking of network elements. Centrality analysis ranks network elements according to their importance within the network structure, and different centrality measures focus on different importance concepts. Central elements of protein networks have been found to be, for example, essential for viability. Highly connected hub genes, largely responsible for maintaining network connectivity, have been found to be much more likely to be essential for yeast survival. However, standard centrality measures may fail to reflect the power of each gene in interaction with the others. Cooperative game theory may also be used to analyse gene expression data. Recently, the class of microarray games has been introduced to quantitatively evaluate the relevance of each gene in regulating or provoking a condition of interest (e.g. a disease), taking into account the observed relationships in all subgroups of genes. In the framework of microarray games, the definition of relevance index for genes is provided in terms of the Shapley value. The higher the number attributed by the Shapley value to a certain gene in a given microarray game, the higher the relevance of that gene for the mechanisms governing the genomic effects of the condition under study. This approach has been successfully applied to real datasets and provides a characterization of a relevance index for genes which is mainly based on the key role played within large partnerships of genes.

Methods

Co-expression networks are connection situations based upon the correlation relationships between individual genes across a gene expression dataset. Nodes are genes and connections are defined by co-expression of two genes. In general, Pearson correlation is the initial measure of gene co-expression. This co-expression measure is then transformed into an adjacency matrix by raising the correlation to a certain power based on a scale free topology criterion. Depending on the aims of the study, weighted networks or un-weighted networks, based on dichotomizing the related correlation matrix, may be considered. In this work, we introduce the class of co-expression network games. A co-expression network game is a cooperative game in characteristic function form (N, v) , where N is the set of genes studied in the expression dataset and v is the characteristic function, which assigns a "worth" to each subset (coalition) of genes in N . The worth of a coalition represents the overall magnitude of its genes' correlation relationships with an a priori selected set of key genes (e.g. a set of genes already known to be involved in a disease), where gene interaction is restricted to the connections of the associated co-expression network. Shapley value of a co-expression network game is considered as gene's relevance.

Results

We introduce a new measure of the importance of genes in a co-expression network, based on cooperative game theoretic model aimed to reflect gene interaction. The new measure, based on the Shapley value for cooperative games, may be used to express the relevance of each gene in interaction with the others. With respect to other statistical methods, cooperative game theory allows for an analysis involving the worth of all possible coalitions, and is suitable for contextualized interpretations using the property driven approach. The main advantage of co-expression network games, with respect previous applications of game theory to gene expression analysis, is a finer resolution of gene interaction investigated in the model, which is based on pair-wise relationships of genes in the network. In addition, this new approach allows for the integration of a priori knowledge concerning the co-expression with key genes, which may be obtained by previous studies. Computational aspects related to the implementation of the relevance index and the design of a model validation study are discussed.

References

- [1] Björn et al. Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics*, 2006, 7:219.
- [2] Carlson et al. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics*, 2006, 7:40
- [3] Moretti et al. Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution. *BMC Bioinformatics*, 2008, 9:361. [4] Moretti et al. The class of microarray games and the relevance index for genes. *Top*, 2007, 15: 256-280.
- [5] Jeong H. et al. Lethality and centrality in protein networks. *Nature*, 2001, 411:41-42.
- [6] Zhang and Horvath. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat Appl Genet Mol Biol*, 2005, 4(1) , Article 17.

Contact : stefano.moretti@istge.it