Heuristics for simulated annealing search of active sub-networks in bio-molecular interaction networks

Simão Soares[1], Isabel Rocha[1], Kiran Raosaheb Patil[2]

[1]IBB-Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, 4710-057 Campus de Gualtar, Braga, Portugal. [2]Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark.

E-mails: simao.soares@deb.uminho.pt, irocha@deb.uminho.pt, krp@bio.dtu.dk

Abstract

Different kinds of 'omics' data for several organisms and bio-molecular interaction networks (e.g. reconstructed networks of biochemical reactions and protein-protein physical interactions) are becoming very common nowadays.

These bio-molecular networks are being used as a platform to integrate genome-scale 'omics' datasets. Identification of sub-networks in these large networks that show maximum collective response to a perturbation is one the interesting problems to solve by using an integrative analysis.

Sub-networks can be hypothesized to represent significant collective biological activity due to the underlying interactions between the bio-molecules. The biological activity can be estimated in several ways- for example coordinated change in the expression level (e.g. mRNA). Identifying these regions reduce complexity of the network to be analyzed in greater detail by revealing the regions that are perturbed by a condition-removing the interactions that are potentially false-positive and not related to the response under study [1].

As the simulated annealing does not guarantee to find the global optimum and may lead to an incomplete picture of the biological phenomenon, we report a method to estimate the theoretical optimal score curve.

The simulated annealing algorithm (SA) used in this study is a slightly modified version of the algorithm by Ideker *et al.* [2]. Each node in the graph is associated with a binary variable turning the node visible or invisible and therefore inducing several sub-graphs. In the standard formulation, the initial solution is obtained by randomly attributing 0 or 1 to the nodes of the graph. Based in concepts described above, we propose an alternative initialization method to improve the performance of the simulated annealing algorithm.

Algorithmic lower bound

Let k* be the size of the complete network that originates the highest value in the maximum score curve. Given a real (non complete) sub-network, it is possible to calculate the algorithmic lower bound by computing the scores of the sub-networks composed of the nodes that exhibit scores higher or equal than the score of the k*th node in the rank of the nodes scores.

Therefore, if a sub-network finding algorithm uses this information during initialization, it is expected to find solutions in the area between the upper bound curve and the algorithmic lower bound.

Simão Soares acknowledges support from "SYSINBIO - Systems Biology as a Driver for Industrial Biotechnology", Coordination and support action (call FP7-KBBE-2007-1).

[1] Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks.2002. Bioinformatics, 18(1):S233-S240

[2] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. 2003. Genome Research, 13(11):2498-504