A transcription network simulator

Di Camillo Barbara[1], Toffolo Gianna[1], Cobelli Claudio[1]

[1]Information engineering department, University of Padova, Padova

Motivation

Studying biological regulatory networks at topological level is a major issue in computational biology studies since this helps understanding the regulation inside the cell. At this purpose, transcription network simulators can be used to generate hypothesis on regulatory mechanisms, to predict cell behaviour and to evaluate reverse engineering algorithms performance on synthetic data, generated from gene networks of known topology. Quality of inference depends on the ability of the simulator to resemble the main features of real regulatory networks, in particular to mimic topology and regulatory mechanisms.

Different models are available in the literature to represent topology of biological networks; however, no one of them is able to describe all the properties observed on real networks: the power-law distribution of connectivity degree, the independence of clustering coefficient on the number of genes in the network, and the small world behaviour. Two main classes of models are available in the literature to describe regulatory mechanisms: differential equation based models and Boolean models of regulation. The former provide continuous data, but in general address simplistic regulatory logic, e.g. additive or multiplicative effects; the latter resemble important aspects of gene regulation such as complex regulatory mechanisms, but do not describe continuous changes in gene expression.

Methods

Here we present a novel transcription network simulator which accounts for main topological and regulatory properties observed in biological networks.

The topology is generated as a scale-free network, by interconnecting sub-network structures which are replicated at different level of network organization (as in fractals). Regulatory sub-networks are generated by randomly assigning a number of regulators to each gene according to a scale-free structure: the probability for each node of having a number of connections with other nodes follows a power-law distribution. The nodes with the highest number of connections are called hubs.

Sub-networks are connected to each other through nodes randomly selected among the hubs. This strategy is iteratively repeated to generate a network characterized by annidated modules, so as to render the simulated network scale-free, but with clustering coefficient not dependent on the number of nodes in the network.

Regulation is modelled using Boolean logic extended to the continuous domain, so as to represent both complex regulatory mechanisms and continuous nature of gene expression data. In particular, transcription regulation is modelled, for each gene x, as a combination of three basic regulatory rules: 1) the function minimum models regulatory effect achieved only if the regulators $r\_1, ..., r\_m$ are simultaneously active: $min(r\_1, ..., r\_m)$; 2) the function maximum models regulatory action which can be alternatively performed by different regulators $r\_1, ..., r\_m$: $max (r\_1, ..., r\_m)$; 3) the minus sign models negative regulation (inhibition or degradation).

For each gene x, the result of the combination of these regulatory rules is a target value T; e.g. $T=min[(-r\_1,max(r\_2,r\_3)]$. To resemble the continuous nature of transcription, the rate of change of gene x expression level at time t is then modelled using differential equation:

$dx(t)/dt = k*(T-x(t))$ where T results from the regulatory rules, and k is a time constant influencing both the rate of transcription and degradation.

Results

Performances of the simulator were considered in terms of generated network topologies and expression profiles. In particular, the simulated topologies were compared to random, scale free and geometric graph; only the new model was able to resemble all the global properties observed in biological networks and to maintain local motif structures, which is also characteristic of real networks.

To test the plausibility of generated profiles, we compared them to a publicly available real data set describing the temporal expression of 102 genes of E. Coli after induction of the expression of a recombinant protein (Schmidt-Heck et al.

2004). Cluster analysis performed on simulated and on real data gave similar results in terms of range and complexity of the profiles. The correlations calculated between any pair of genes had similar distribution on simulated and real data. The same result were obtained using the pair-wise partial correlation and covariance.

The simulator, by integrating two main features of transcription networks such as topological and regulatory properties, should provide a reliable test bed for reverse engineering algorithms.

Email: dicamill@dei.unipd.it