

SCALE-FREE STRUCTURE AND TOPOLOGICAL PROPERTIES IN REVERSE ENGINEERING OF GENE REGULATORY NETWORKS.

SILVANA BADALONI, MARCO FALDA, FRANCESCO SAMBO

Department of Information Engineering, University of Padova, via Gradenigo 6/b, 35100 Padova, Italy.

Extended abstract

1. Introduction

A wide range of natural, technological and social systems, such as gene regulatory networks, the Internet, electrical power grids and biological signaling networks in the cell, have been recently proven to show scale-free network structure. The distinct feature of scale-free networks is that node degree distribution follows a power law [2], exhibiting few highly connected nodes (*hubs*) and the majority of nodes with low connectivity.

Up to now, the scientific community has focused mostly on verifying the presence or absence of this kind of organization in the most scientifically interesting networks; less attention has been given to the design of algorithms for exploiting information on this particular topological structure in solving problems related to such networks.

In this paper, we analyze the structure of a specific class of network that exhibit the scale free characteristic, Gene Regulatory Networks, in order to address the problem of reverse engineering from DNA-microarray data, *i.e.* the reconstruction of causal relations among genes starting from their expression profiles. The problem is inherently difficult, because of the extremely wide amount of data to be processed, and topological information can be useful to reduce the size of the search space.

Reverse Engineering methods on real datasets and realistically simulated data are known to exhibit poor performance [1]. In this work, we investigate how performance depends on network structure, to understand if there exist parts of the network easier to be identified and, on the contrary, sections intrinsically difficult to be reconstructed. To this purpose, two Reverse Engineering algorithms were tested on simulated data: Dynamic Bayesian Networks (DBN)

[4], which works with continuous data, and Causal Networks (CNET) [5,6], which works with quantized data. Resulting networks were inspected, searching for regions identified by both algorithms or, on the opposite, missed by both. In particular, we explored how performance depends on local network structure.

2. Methods

We randomly generated 15 networks of 12 genes with the simulator described in [2]. This simulator generates network topologies by recursively assembling motif structures characteristic of biological networks, using an iterative strategy that guarantees scale-free degree distribution.

In each network, the hub with the highest out-degree was then externally stimulated with three kinds of stimuli: a sinus, a ramp and a step, so to represent a variety of possible experimental situations. We then obtained the three different datasets, on which we tested both DBN and CNET algorithms.

We first analyzed how much the behaviours of the two algorithms overlap, *i. e.* if there exist portions of the networks intrinsically simple or difficult for both: we considered, for every edge, the number of times it was identified by each algorithm in each dataset. The results were inspected, searching for common behaviour patterns.

3. Results

No stimulus, among sinus, ramp and step, allows obtaining significantly higher performance in all the networks (or in the majority of them) and no algorithm, between DBN and CNET, performs systematically better than the other.

Performances depend on the distance between the stimulated hub and the reconstructed edge, being significantly higher on edges outgoing from the hub and decreasing with the distance. A similar result has been observed also in [1], with different simulated data and a differential equation based Reverse Engineering model.

Moreover, the behaviour of the two algorithms tends to be homogenous on particular motifs, which are recognised with significantly higher performance than others; for example, circles of genes regulating each other in a chain are recognized with significantly higher frequency [6] (see fig. 1).

In addition, performance on edges ingoing or outgoing from a node is significantly better for nodes with lower clustering coefficient, *i. e.* nodes with fewer connections among their neighbours.

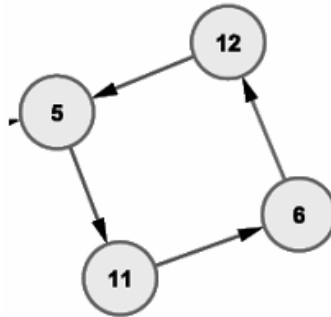


Figure 1. Network motif easy to identify by DBN and CNET.

4. Conclusions and future directions

These results will be useful in the design of new algorithms, which will exploit the increase in performance near external stimuli, the information on common network motifs and, in general, the scale-free property of biological networks.

5. Acknowledgements

This work is carried over in collaboration with Barbara Di Camillo and Gianna Toffolo of the Bioengineering Research Group of the University of Padova.

References

1. Bansal, M., Belcastro, V., Ambesi-Impiombato, A., Di Bernardo, D.: How to infer gene networks from expression profiles. *Molecular Systems Biology* 3:78 (2007)
2. Barabasi, A.L., Albert R., Emergence of scaling in random networks. *Science*, 286(5439):509–512, (1999)
3. Di Camillo, B., Toffolo, G., Cobelli, C.: A Gene Network Simulator to Assess Reverse Engineering Algorithms. *Annals Of The New York Academy Of Sciences*. ISSN:0077-8923. Accepted. (2008)
4. Ferrazzi, F., Sebastiani, P., Ramoni, M.F., Bellazzi, R.: Bayesian approaches to reverse engineer cellular systems: a simulation study on nonlinear Gaussian networks. *BMC Bioinformatics* 8(Suppl 5):S2 (2007)
5. Sambo, F., Di Camillo, B., Toffolo, G.: CNET: an algorithm for Reverse Engineering of Causal Gene Networks. In: *NETTAB08*, Milan, Italy (2008)
6. Sambo, F., Di Camillo, B., Toffolo, G.: Role of network structure and experimental design on the performance of two Reverse Engineering methods. In: *European Conference on Computational Biology 08*, Cagliari, Italy (2008)