

NETWORK INFERENCE BY INTEGRATING BICLUSTERING AND FEATURE SELECTION

Robrecht Cannoodt^{1,2,3,4,*}, *Joeri Ruysinck*⁴, *Katleen De Preter*³, *Tom Dhaene*⁴, *Yvan Saeys*^{1,2}
 VIB Inflammation Research Center, Ghent, Belgium¹; Department of Biomedical Molecular Biology, Center for Medical Genetics, Ghent, Belgium²; Department of Pediatrics and Genetics, Ghent University, Ghent, Belgium³; Department of Information Technology, Ghent University – iMinds, Ghent, Belgium⁴. *robrecht.cannoodt@ugent.be

In order to develop better therapies to combat specific abnormalities present in the gene regulatory network (GRN) of cancer patients, it is crucial to gain a better understanding of regulatory networks in complex biological systems. An important class of methods in systems biology are network inference (NI) methods, which aim to reconstruct a GRN from high-throughput data (e.g. microarrays or next-generation sequencing).

INTRODUCTION

GENIE3¹ is a state-of-the-art method which employs feature selection to identify the best subset of regulators for each gene. While this method is amongst the best performing NI methods, it fails to take into account expected topological properties of a GRN: a GRN consists of modules, each of which consists of genes coregulated by a common set of regulators.

METHODS

We present BiGENIE, a method which takes the modular topology of a GRN into account. Figure 1 shows the difference in process between GENIE3 and BiGENIE. By firstly inferring modules – groups of genes coregulated by a common regulator – using several biclustering methods, the overall topology of the network is reconstructed. Subsequently, the regulator genes for each of the modules is inferred using GENIE3.

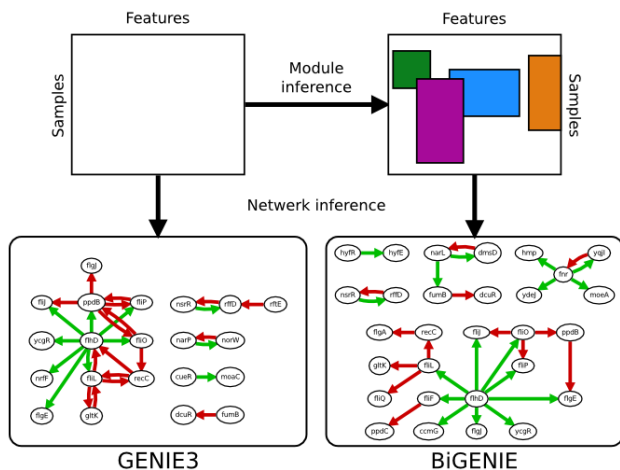


FIGURE 1. A comparison between GENIE3 and BiGENIE in terms of structure and the resulting inferred network. Green arrows represent correctly predicted interactions, while red arrows represent falsely predicted interactions.

RESULTS & DISCUSSION

The Area-Under-ROC (AUROC) and Area-Under-Precision-Recall (AUPR) values are commonly used metrics to objectively evaluate the performance of NI methods. Table 1 contains AUROC and AUPR values of the GENIE3 and BiGENIE methods evaluated on 11 different datasets. The datasets consist of five small in silico networks from the DREAM4^{2,3,4} competition, one large in vitro dataset from the DREAM5⁵ competition, and five moderately sized networks generated by the GeneNetWeaver⁶ tool (using the default settings).

Method	GNW200		DREAM4		DREAM5	
	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
BiGENIE	0.762	0.171	0.771	0.148	0.642	0.045
GENIE3	0.722	0.084	0.767	0.187	0.618	0.094
# genes	200		100		4511	
# samples	200		100		805	

TABLE 1. Evaluation of the BiGENIE and GENIE3 methods.

From these results can be concluded that BiGENIE works very well on the GNW200 datasets. However, there is still room for improvement when BiGENIE is applied to networks of a different magnitude, as its AUPR values were less than GENIE3. By manually inspecting the results we noticed BiGENIE and GENIE3 do produce complementary results, so that these results could be combined to produce a better solution.

REFERENCES

- Vân Anh Huynh-Thu *et al.* “Inferring Regulatory Networks from Expression Data Using Tree-Based Methods”. In: PLoS ONE 5.9 (Sept. 2010), e12776.
- Daniel Marbach *et al.* “Generating Realistic In Silico Gene Networks for Performance Assessment of Reverse Engineering Methods”. In: Journal of Computational Biology 16.2 (2009), pp. 229–239.
- Daniel Marbach *et al.* “Revealing strengths and weaknesses of methods for gene network inference”. In: Proceedings of the National Academy of Sciences 107.14 (2010), pp. 6286–6291.
- Robert J. Prill *et al.* “Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges”. In: PLoS ONE 5.2 (Feb. 2010), e9202.
- Marbach Daniel *et al.* “Wisdom of crowds for robust gene network inference”. In: Nat Meth 9.8 (Aug. 2012), pp. 796–804.
- Thomas Schaffter *et al.* “GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods.” In: Bioinformatics 27.16 (2011), pp. 2263–2270.