

Talk

Ensemble and Greedy Approach for the Reconstruction of Large Gene Co-Expression Networks



Francisco Gómez-Vela (1), Fernando M. Delgado-Chaves (2), Domingo S. Rodríguez-Baena (1), Miguel García-Torres (1) and Federico Divina (1).

(1) Computer Science Division, Pablo de Olavide University, ES-41013 Seville, Spain

(2) Faculty of Experimental Sciences, Pablo de Olavide University, ES-41013 Seville, Spain

Keywords: gene networks; scale-free topology; ensemble networks; graph theory; computational biology; co-expression networks; biomarkers discovery

ABSTRACT

In the recent years, the vast amount of genetic information generated by high-throughput approaches, have led to the need of new methods for data handling. The integrative analysis of diverse-nature gene information could provide a much-sought overview to study complex biological systems and processes. In this sense, Co-expression Gene Networks (CGN) have become a powerful tool in the comprehensive analysis of gene expression. Such networks represent relationships between genes (or gene products) by means of a graph composed of nodes and edges, where nodes represent genes and edges the relationships among them. Amongst the main features of CGN, sparseness and scale-free topology may notably affect the latter network analysis. Within this framework, structure optimization techniques are also important in the reduction of the size of the networks, not only improving their topology but also keeping a positive prediction ratio. On the other hand, ensemble strategies have significantly improved the precision of results by combining different measures or methods.

In this work, we present Ensemble and Greedy networks (EnGNet), a novel two-step method for CGN inference. First, EnGNet uses an ensemble strategy for co-expression networks generation. Final score is estimated by major voting among three different methods, i.e. Spearman and Kendall coefficients and Normalized Mutual Information. Second, a greedy algorithm optimizes both the size and the topological features of the network. Not only do achieved results show that this method is able to obtain reliable networks, but also that it significantly improves the topology of the networks.

Moreover, the usefulness of the method is proven by an application to a human dataset on post-traumatic stress disorder (PTSD), revealing an innate immunity-mediated response to this pathology in accordance with previous studies. These results are indicative of the potential of CGN, and EnGNet in particular, in the unveiling of the genetic causes for complex diseases. Finally, the implications of CGN in biomarkers discovery, could lead research towards earlier detection and effective treatment of these diseases.

REFERENCES

- Gómez-Vela, F., Delgado-Chaves, F. M., Rodríguez-Baena, D. S., García-Torres, M., & Divina, F. (2019). Ensemble and Greedy Approach for the Reconstruction of Large Gene Co-Expression Networks. *Entropy*, 21(12), 1139.
- Delgado, F. M., & Gómez-Vela, F. (2019). Computational methods for Gene Regulatory Networks reconstruction and analysis: A review. *Artificial intelligence in medicine*, 95, 133-145.
- Gómez-Vela, F., Rodríguez-Baena, D. S., & Vázquez-Noguera, J. L. (2018). Structure optimization for large gene networks based on greedy strategy. *Computational and mathematical methods in medicine*, 2018.

