

Candidature proposal

Miguel Romero

July 6, 2020

P5 - ÓMICAS

Doctorado en Ingeniería

Pontificia Universidad Javeriana de Cali

Table of contents

1. Previous work:
Node label and link prediction
2. Proposal

Previous work:

Node label and link prediction

Objective

Prediction of gene functions based on biological data.

How?

- gene co-expression network
- existing knowledge body of gene functions of a given genome
- supervised machine learning and probabilistic model

Approaches

1. Machine learning + structural properties of gene co-expression + function prediction (independently) [2],
2. HBN probabilistic model + gene co-expression network + function hierarchical structure + function ancestral relation,
3. Machine learning + gene co-expression network + function hierarchical structure + function ancestral relation + structural properties of gene co-expression + representations of nodes (node2vec).

Objective

Prediction of links in dynamic networks.

How?

- spectral analysis (eigen decomposition)
- networks grow over time
- formation of new edges can be expressed as transformations of the spectrum (eigenvalues)

Approaches

1. Spectral analysis + multi-graphs (Twitter networks) [3],
2. Spectral analysis + simple graphs (Twitter networks) + Rayleigh quotient + learning models (spectrum),

Proposal

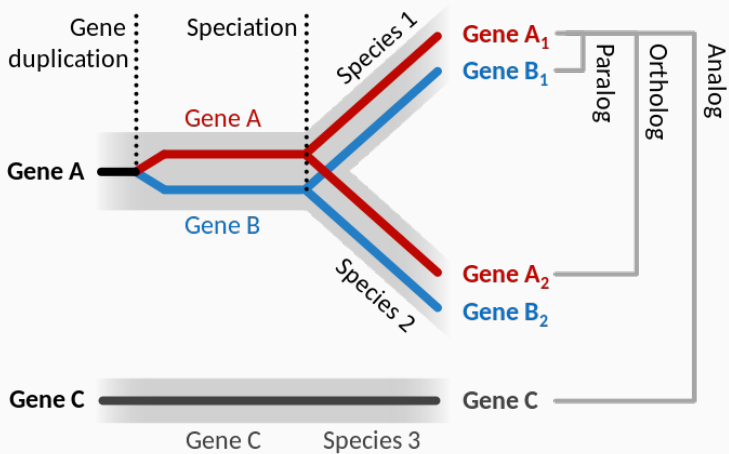
- Identifying functions of genes as an important step to analyze and understand how the genome of an specific organism works.
- Large amount of genes and gene products with no known functional relation identified, **specially for some organism.**

Issues

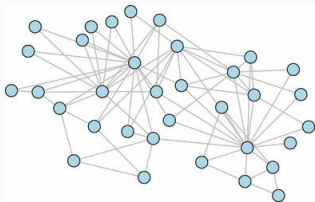
- Lack of biological data *for some organism*,
- Unstructured data,
- New data comes from biological experiments.

- Some organism (model organisms, e.g., arabidopsis) have been widely studied (in vivo and in silico), and there is a comprehensive knowledge body of gene functions for them.
- It is possible to map genes from different organism using **orthology** (e.g., Cenicaña presentation).

Ideas: Orthology

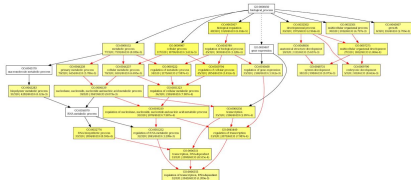


Co-expression Network
(simple, undirected, unweighted)

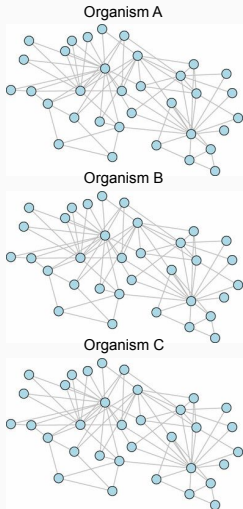


+

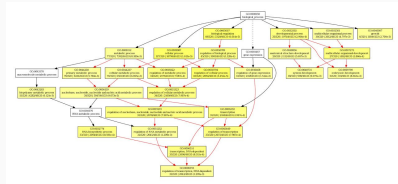
GO hierarchy
(directed acyclic graph)



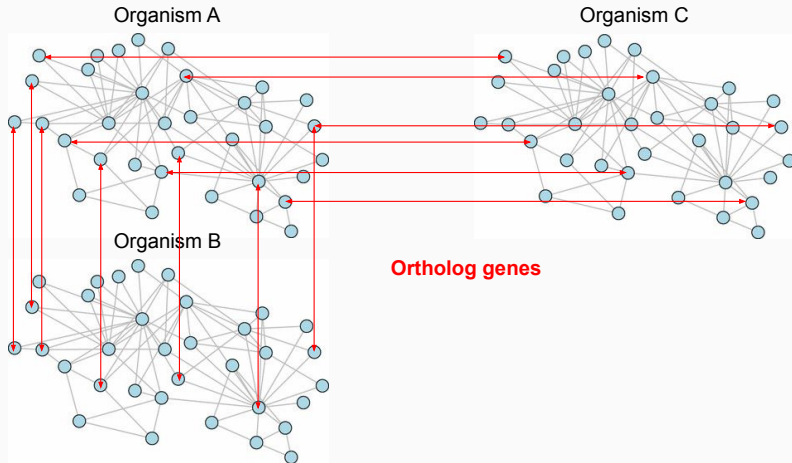
Co-expression Network
(simple, undirected, unweighted)



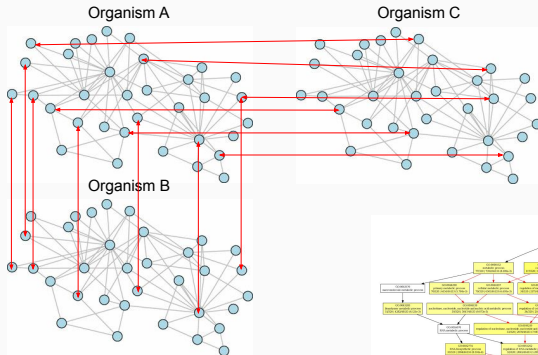
GO hierarchy
(directed acyclic graph)



Co-expression Network (simple, undirected, unweighted)



Ideas

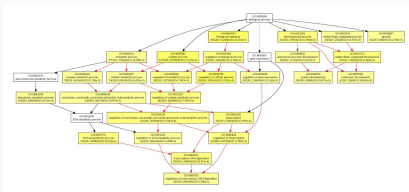


Co-expression Network
(simple, undirected, unweighted)

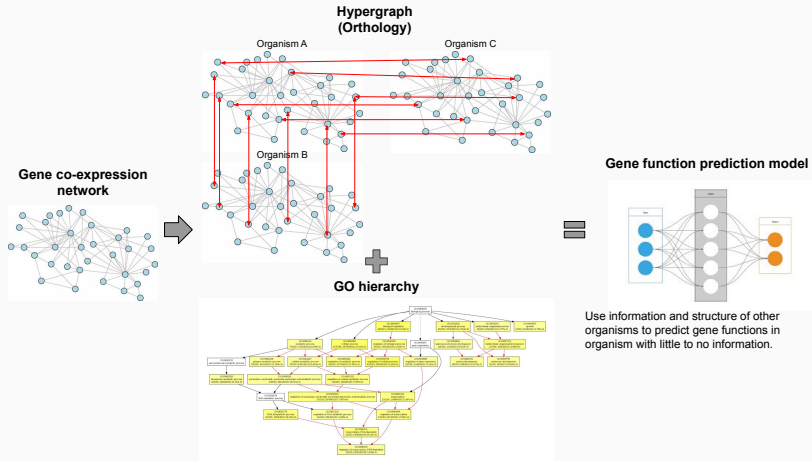
Ortholog genes



GO hierarchy
(directed acyclic graph)





Ideas



- Propose a hypergraph node label predictive model (probabilistic or machine learning) that uses information from related networks,
- Use spectral analysis, e.g., through *Spectral Network Embedding* [4],
- Use GO hierarchical structure (*true-path rule* consistency),
- Use network structural properties.

Questions?

-  X. Jiang, N. Nariai, M. Steffen, S. Kasif, and E. D. Kolaczyk.
Integration of relational and hierarchical network information for protein function prediction.
BMC Bioinformatics, 9(1):350, 2008.
-  M. Romero, J. Finke, M. Quimbaya, and C. Rocha.
In-silico Gene Annotation Prediction Using the Co-expression Network Structure.
In H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, and L. M. Rocha, editors, *Complex Networks and Their Applications VIII*, volume 882, pages 802–812. Springer International Publishing, Cham, 2020.



M. Romero, C. Rocha, and J. Finke.

Spectral Evolution of Twitter Mention Networks.

In H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, and L. M. Rocha, editors, *Complex Networks and Their Applications VIII*, volume 881, pages 532–542. Springer International Publishing, Cham, 2020.



J. Zhang, Y. Wang, J. Tang, and M. Ding.

Spectral Network Embedding: A Fast and Scalable Method via Sparsity.

arXiv:1806.02623 [cs], June 2018.

Thanks!