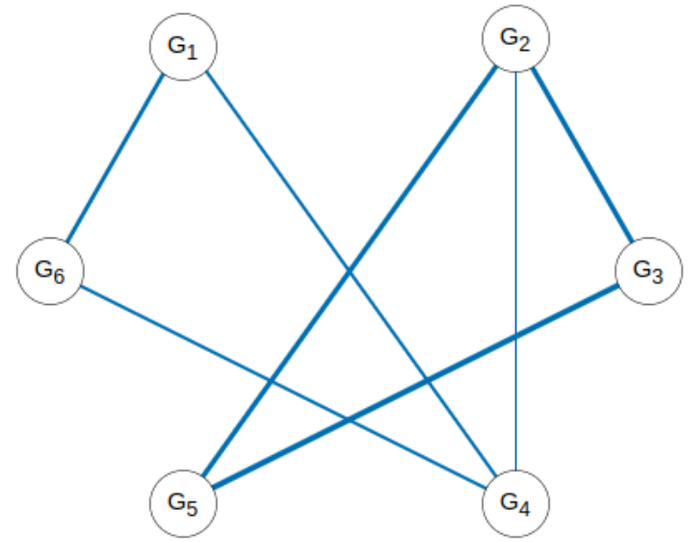
# ÓMICAS

Julian Colorado Montaño
Investigador Principal Proyecto Ómicas
colorado@javeriana.edu.co

El futuro es de todos
Gobierno de Colombia
COLOMBIA CIENTÍFICA
Conocimiento Global para el Desarrollo

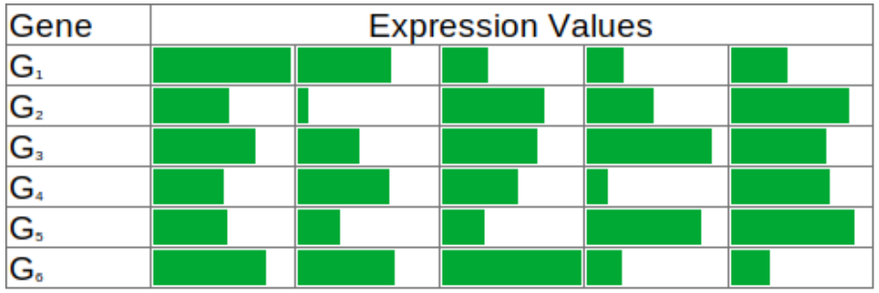
# Characterization of correlation measures with a novel network-based threshold for generating gene co-expression networks. Case study: Rice

Nicolás López-Rozo, Miguel Romero, Jorge Finke, Camilo Rocha

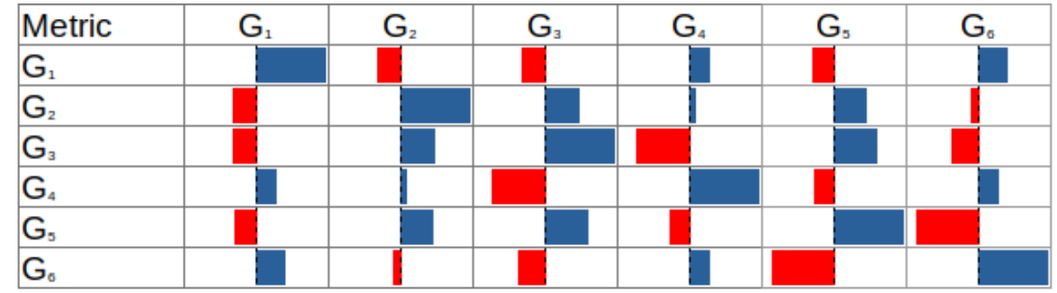



ÓMICAS program – Project P5: In-silico

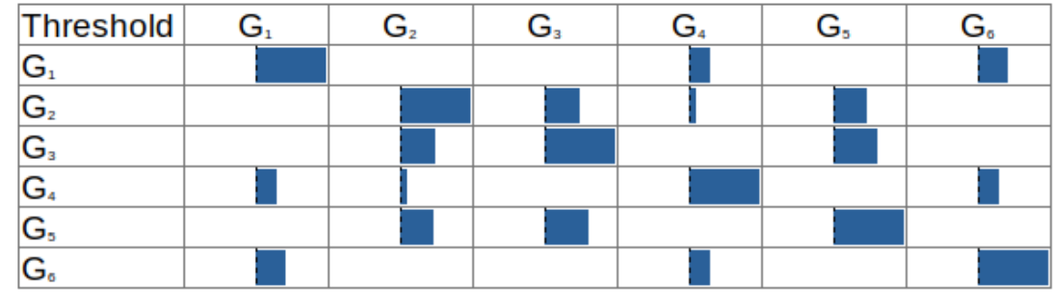
# MOTIVATION: ASSESS CRITICAL STEPS IN GENE ENRICHMENT ANALYSIS



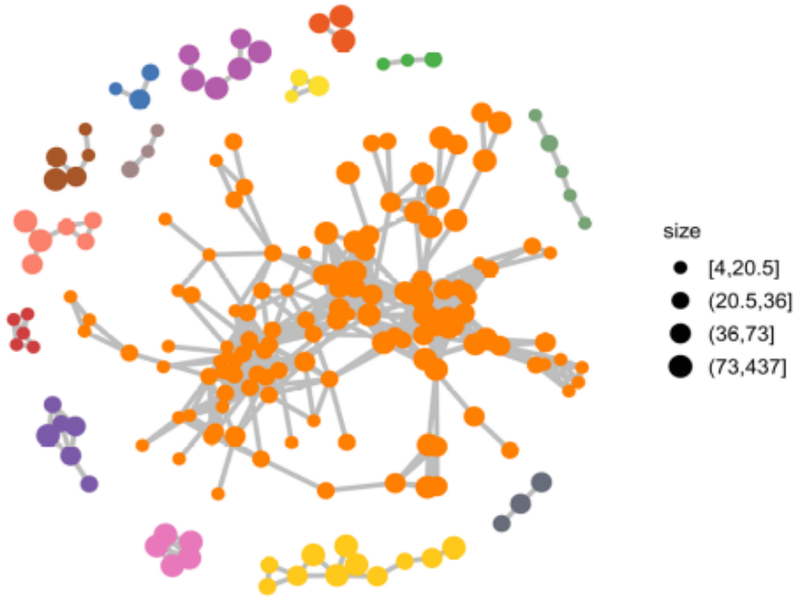

Metric Computation




Threshold Procedure

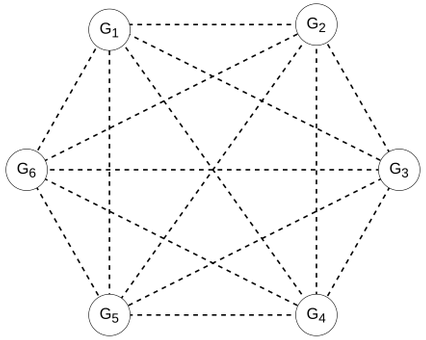


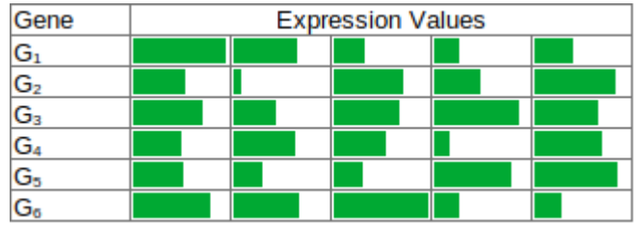

Gene Enrichment



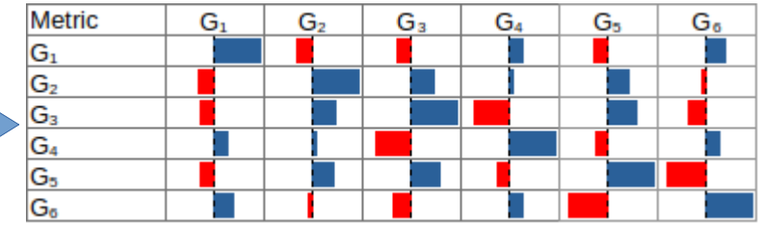

1. Pearson limitations.
2. Threshold is arbitrary.

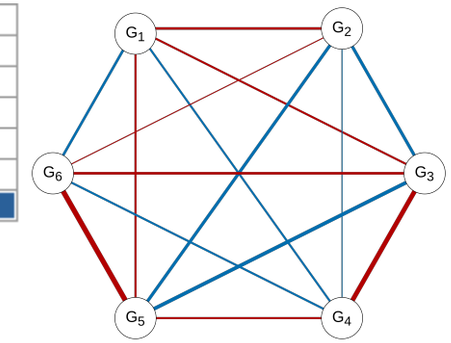
# WORKFLOW: TOWARDS FUNCTION PREDICTION



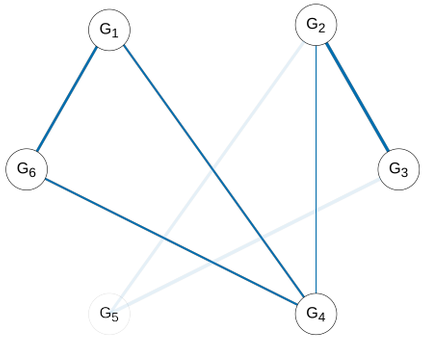



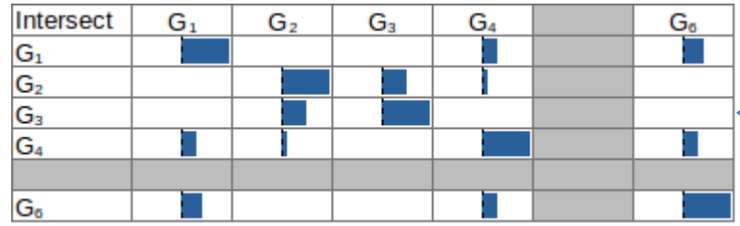

PCC
BICOR
DCORR
MIC
TIC
RIC

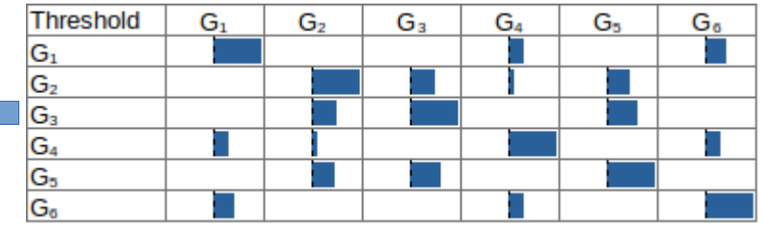


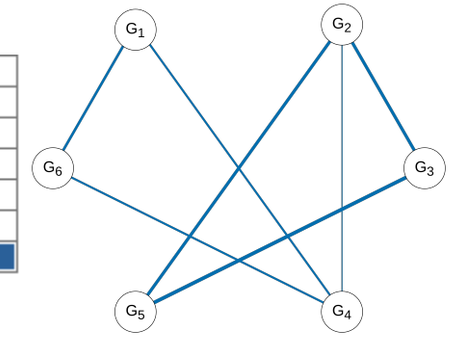



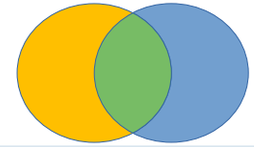

Compute
Topological
Threshold



# DATA

Expression data for *Oryza sativa Japonica:*

- Number of genes: 23374
- Number of samples: 2678
- Data source: NCBI GEO
- Expression value range: 0.009 – 280954

For the classification models:

- 5-fold cross-validation
- 50 repetitions for each model

Selected correlation metrics:

- Pearson's Correlation Coeficient (PCC)
- Biweight Midcorrelation (BICOR)
- Distance Correlation (DCORR)
- Maximal Information Coefficient (MIC)
- Total Information Coefficient (TIC)
- Randomized Information Coefficient (RIC)

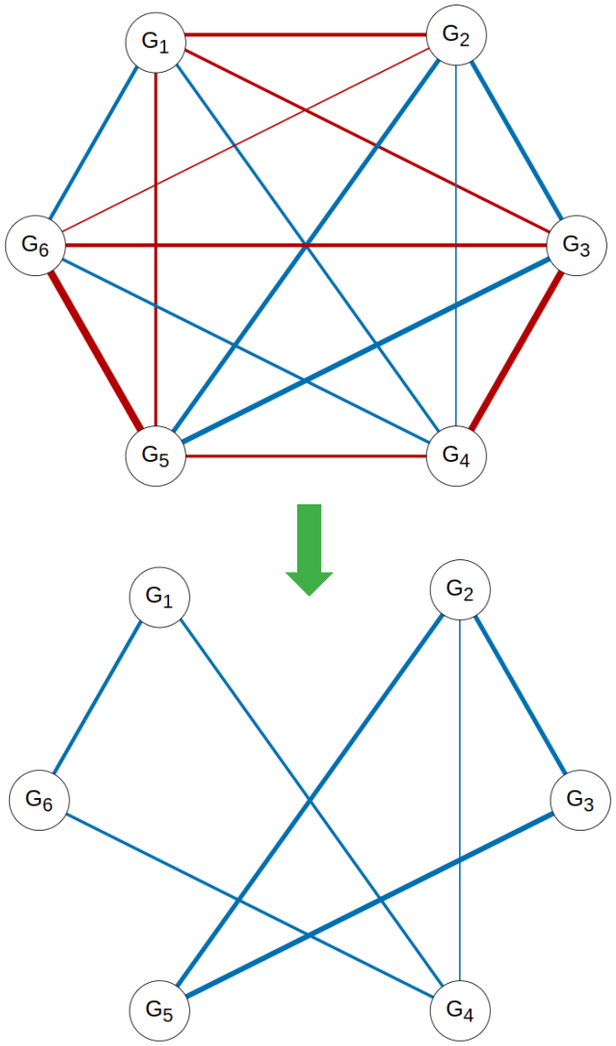
$$\mathrm{PCC}(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$\mathrm{BICOR}(X,Y) = \sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i,$$

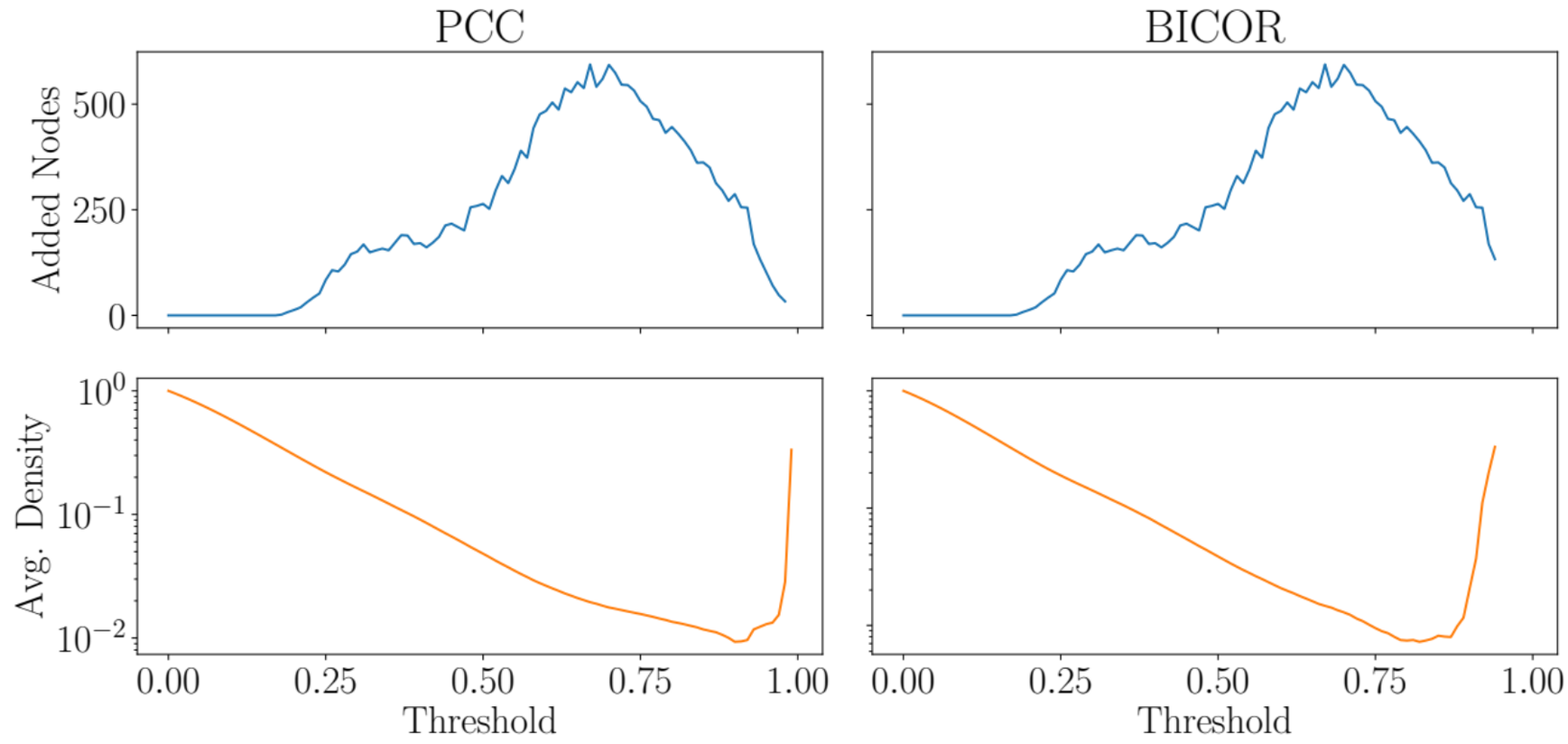
$$\tilde{x}_i = \frac{(x_i - \mathrm{med}(x))w_i^{(x)}}{\sqrt{\sum_{j=1}^{n}[(x_j - \mathrm{med}(x))w_j^{(x)}]^2}}$$

$$\mathrm{DCORR}(X,Y) = \frac{\mathrm{DCov}^2(X,Y)}{\sqrt{\mathrm{DCov}^2(X,X)\,\mathrm{DCov}^2(Y,Y)}}$$

# PROPOSED THRESHOLD COMPUTATION, BASED ON THE TOPOLOGY




$$t_{\mathbf{C}} = \min\left(\text{argmax}(|\Delta V_{I,\mathbf{C}}|), \text{argmin}(\bar{\rho}_{\mathbf{C}})\right)$$

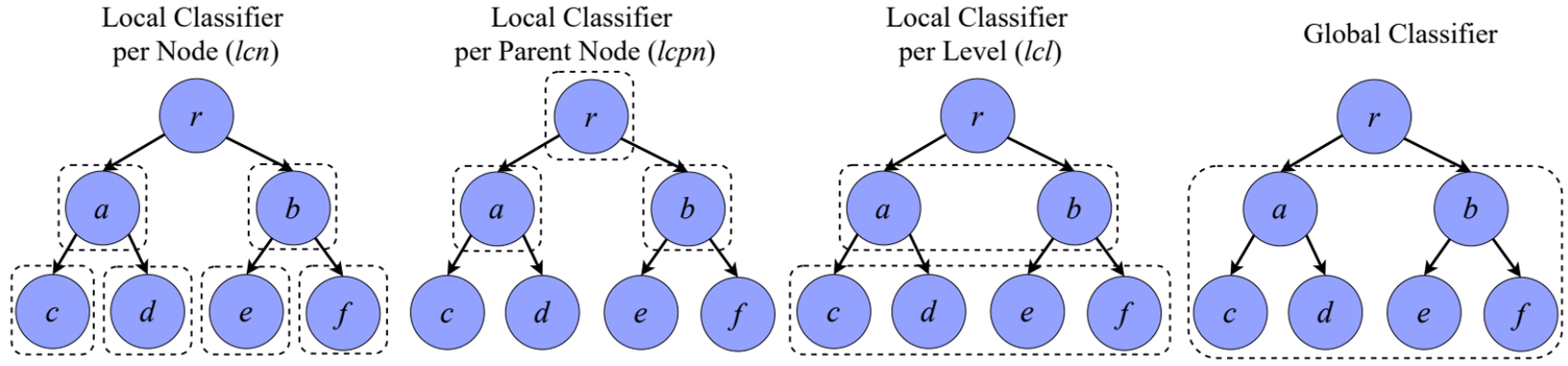


# PERFORMANCE ASSESSMENT: GENE FUNCTION PREDICTION

Four classification models taking advantage of the hierarchical organization of annotations:

- Local Classifier per Node (lcn)
- Local Classifier per Level (lcl)
- Local Classifier per Parent Node (lcpn)
- Global (multi-label) Classifier

Evaluation metric for comparison: Area under the average precision-recall curve



# PERFORMANCE RESULTS (1): COMPARISON OF CORRELATION METRICS

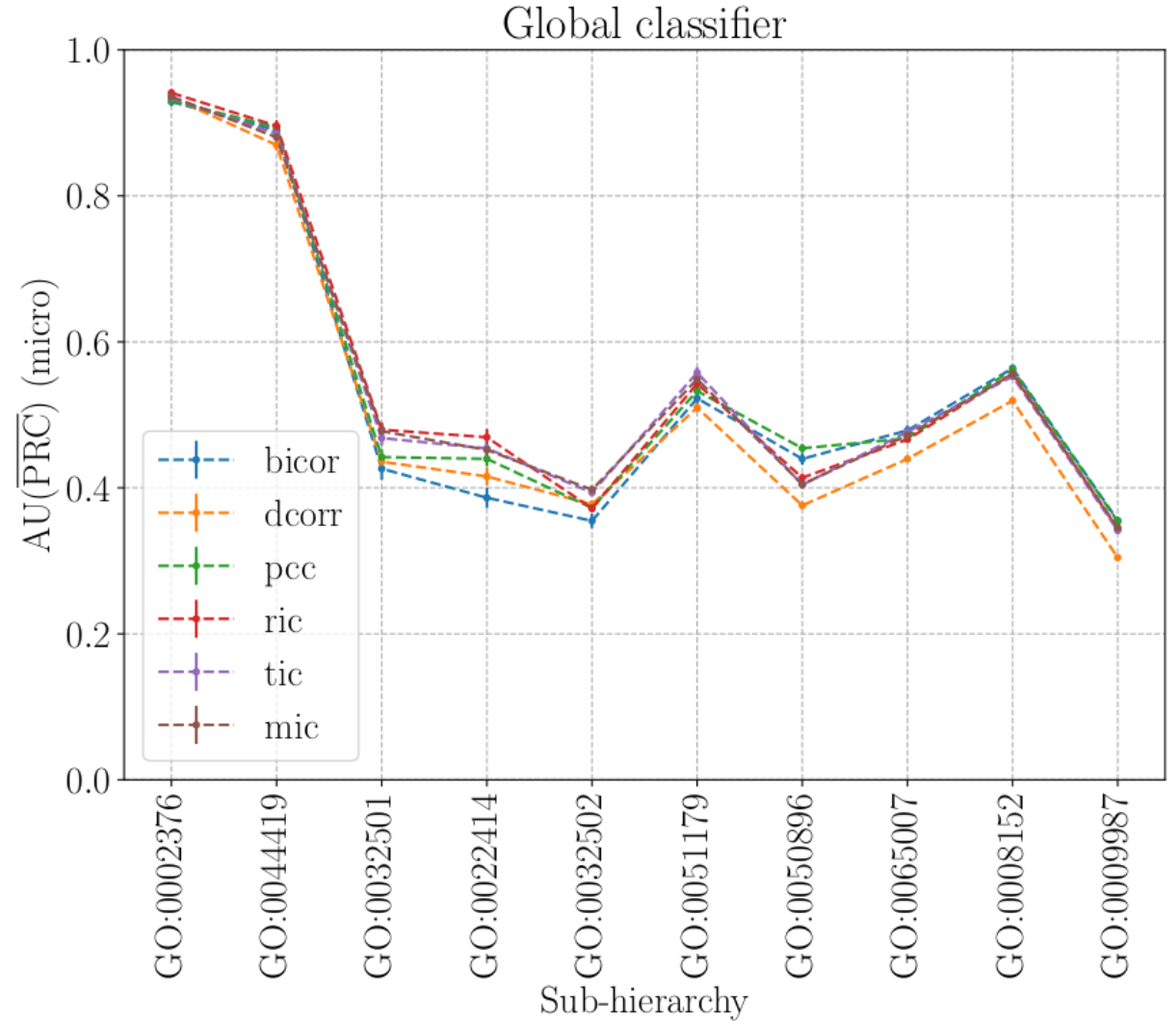
PROPERTIES OF THE RESULTING NETWORKS

| Network | Nodes | Edges | Components | Avg. degree |
|---|---|---|---|---|
| PCC | 11,241 | 1,195,905 | 473 | 212.78 |
| BICOR | 7,344 | 382,202 | 586 | 104.09 |
| DCORR | 13,069 | 2,184,554 | 12 | 334.31 |
| MIC | 14,233 | 5,790,949 | 255 | 813.74 |
| TIC | 13,549 | 5,276,496 | 275 | 778.88 |
| RIC | 13,383 | 3,261,088 | 314 | 487.35 |

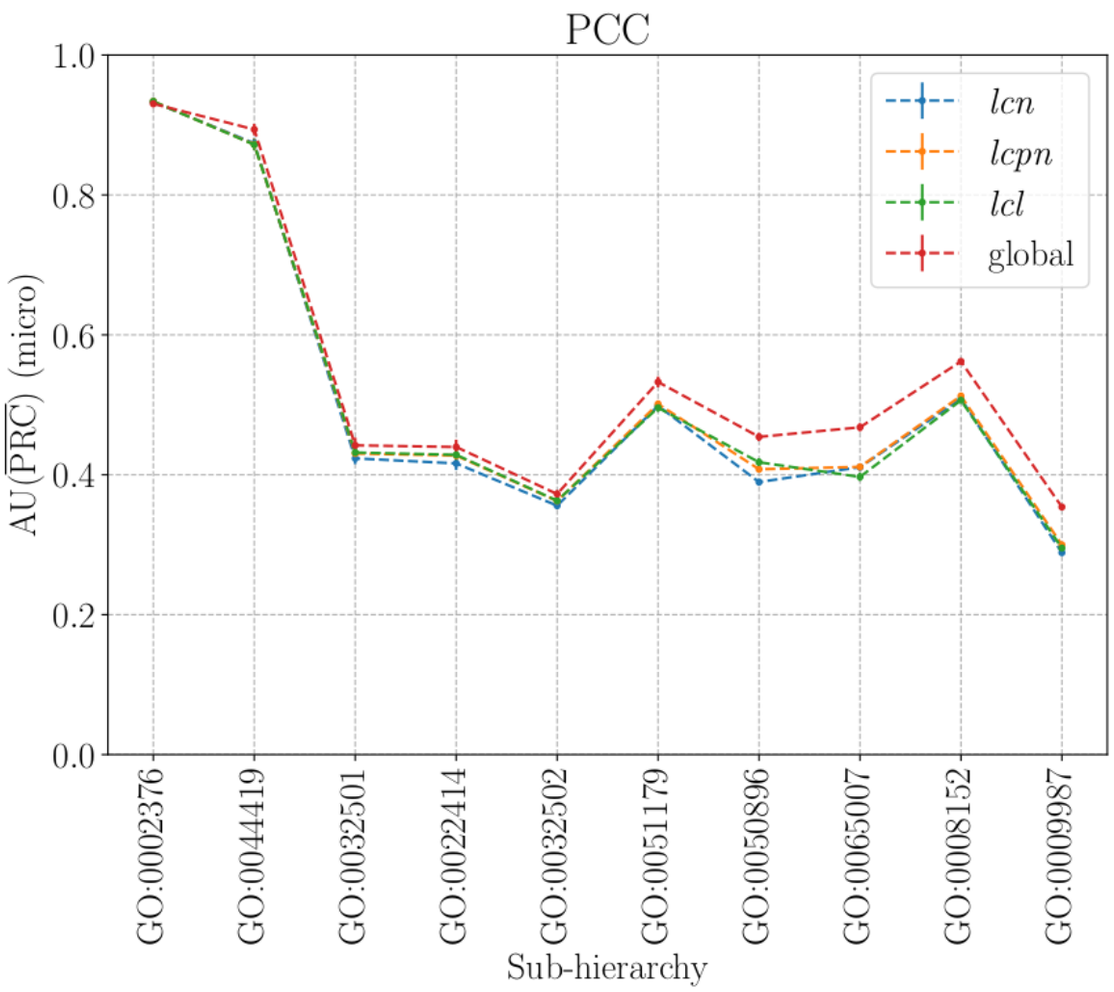
SUB-HIERARCHIES OF BIOLOGICAL PROCESSES[a]

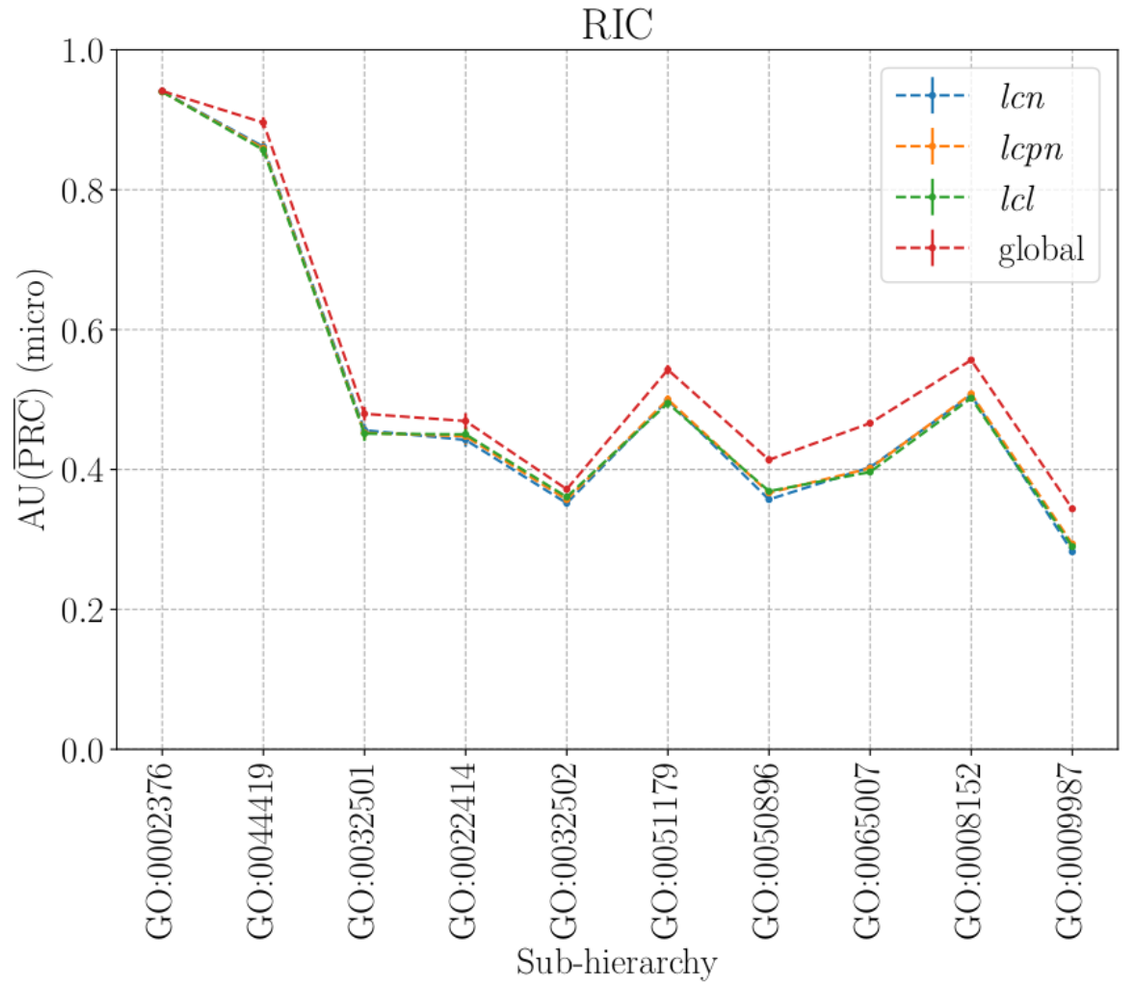
| Hierarchy[b] | Description | Functions |
|---|---|---|
| GO:0002376 | Immune system process | 9 |
| GO:0044419 | Biological process involved in inter-species interaction between organisms | 13 |
| GO:0032501 | Multicellular organismal process | 18 |
| GO:0022414 | Reproductive process | 24 |
| GO:0032502 | Developmental process | 53 |
| GO:0051179 | Localization | 86 |
| GO:0050896 | Response to stimulus | 102 |
| GO:0065007 | Biological regulation | 184 |
| GO:0008152 | Metabolic process | 399 |
| GO:0009987 | Cellular process | 517 |

[a] Each sub-hierarchy is considered an independent dataset.
[b] Each sub-hierarchy is represented by its root.



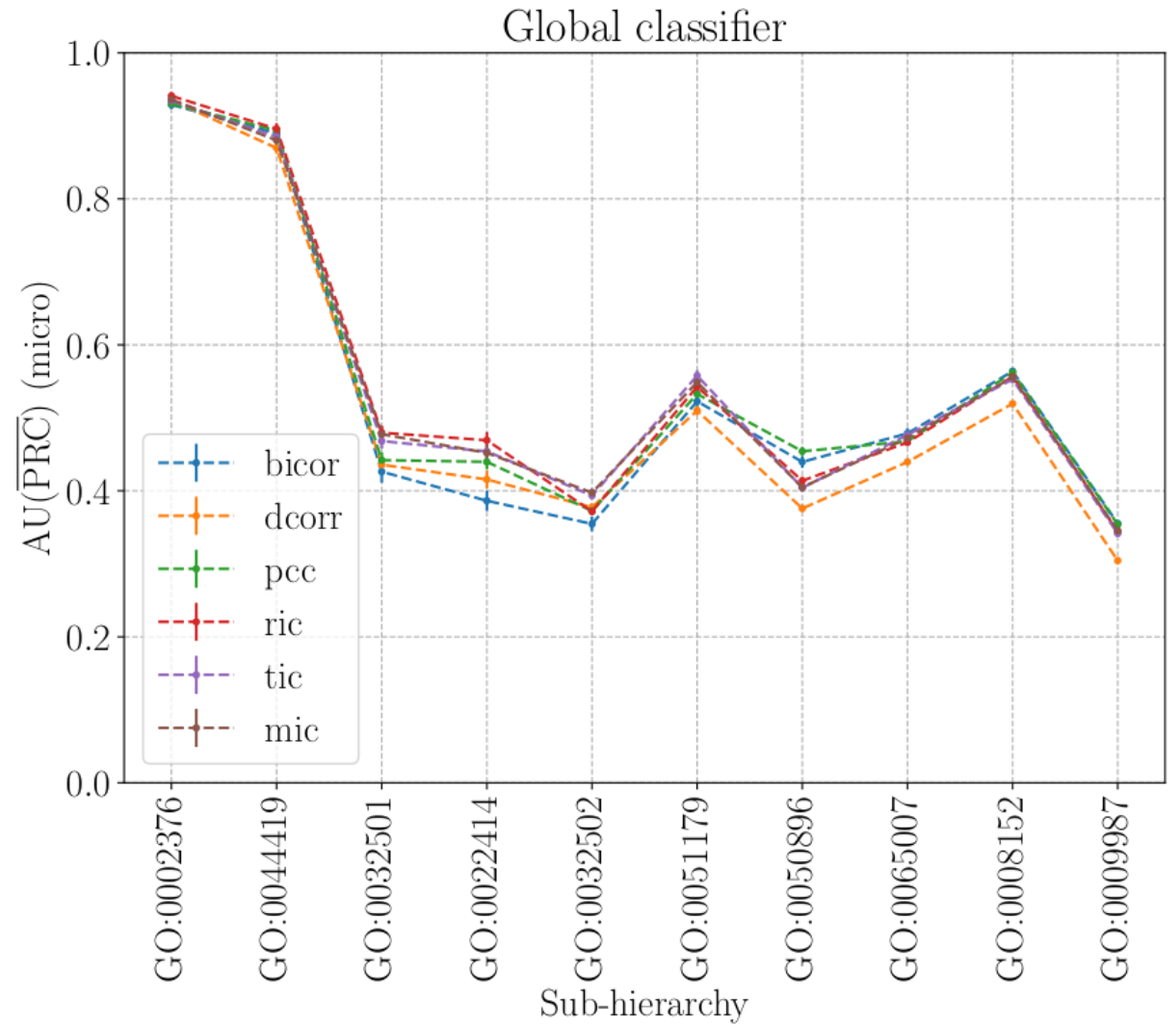
# PERFORMANCE RESULTS (2): COMPARISON OF CLASSIFICATION MODELS



# CONCLUSIONS

- Global multi-label classification model performs better with respect to AU-PRC
- GO hierarchies with less than 25 terms have a better performance with RIC GCN
- Similarly, hierarchies with more than 180 terms get better results with BICOR GCN
- Future work: Explore different threshold approaches and characterize network growth as a function of the cut-off value



# ÓMICAS

## Aliados

CIAT
International Center for Tropical Agriculture
Since 1967 Science to cultivate change

cenicaña

FEDEARROZ
FEDERACIÓN NACIONAL DE ARROCEROS

NIAB

intelecto
Soluciones y Tecnología

HITECH
AUTOMATIZACION S.A.

Caltech

GHENT UNIVERSITY

東京大学
THE UNIVERSITY OF TOKYO

ILLINOIS

Pontificia Universidad JAVERIANA Cali
IES Ancla

UNIVERSIDAD DEL QUINDÍO

Universidad de Ibagué

UNILLANOS
UNIVERSIDAD DE LOS LLANOS

Universidad de los Andes
Colombia

UNIVERSIDAD ICESI

Pontificia Universidad JAVERIANA Bogotá



## Apoyan

El futuro es de todos
Gobierno de Colombia

COLOMBIA CIENTÍFICA
Conocimiento Global para el Desarrollo

# ÓMICAS

Julian Colorado Montaño
Investigador Principal Proyecto Ómicas
colorado@javeriana.edu.co

El futuro es de todos
Gobierno de Colombia
COLOMBIA CIENTÍFICA
Conocimiento Global para el Desarrollo