

ómica

A stylized graphic on the right side of the word 'ómica'. It features a white DNA double helix structure. The top part of the helix is composed of a series of white dots of varying sizes, arranged in a spiral pattern. The bottom part of the helix is composed of a series of white lines, also arranged in a spiral pattern. The entire graphic is set against a solid green background.

El futuro
es de todos

Gobierno
de Colombia



OVERLAPPING MODULES DETECTION TECHNIQUES

Camila Riccio Rengifo

In-silico

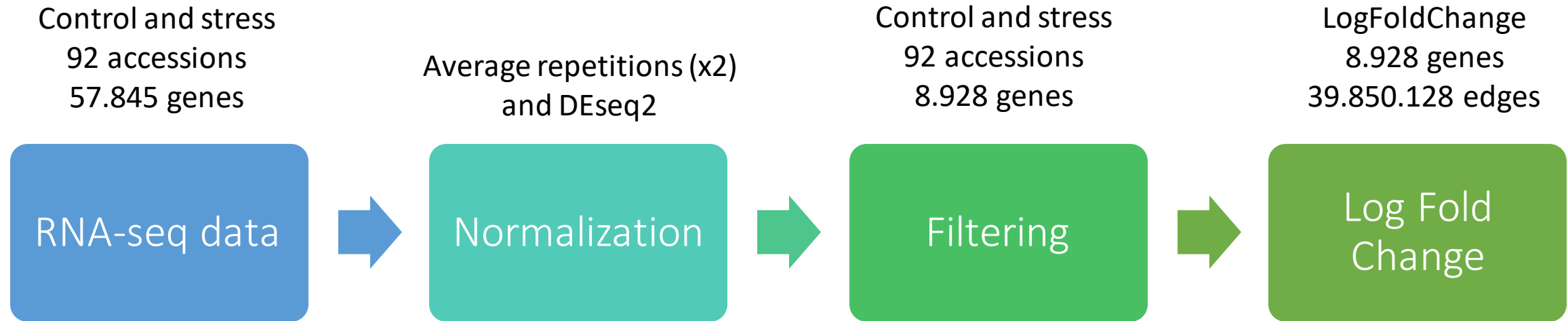


El futuro
es de todos

Gobierno
de Colombia



Data preprocessing



In-silico



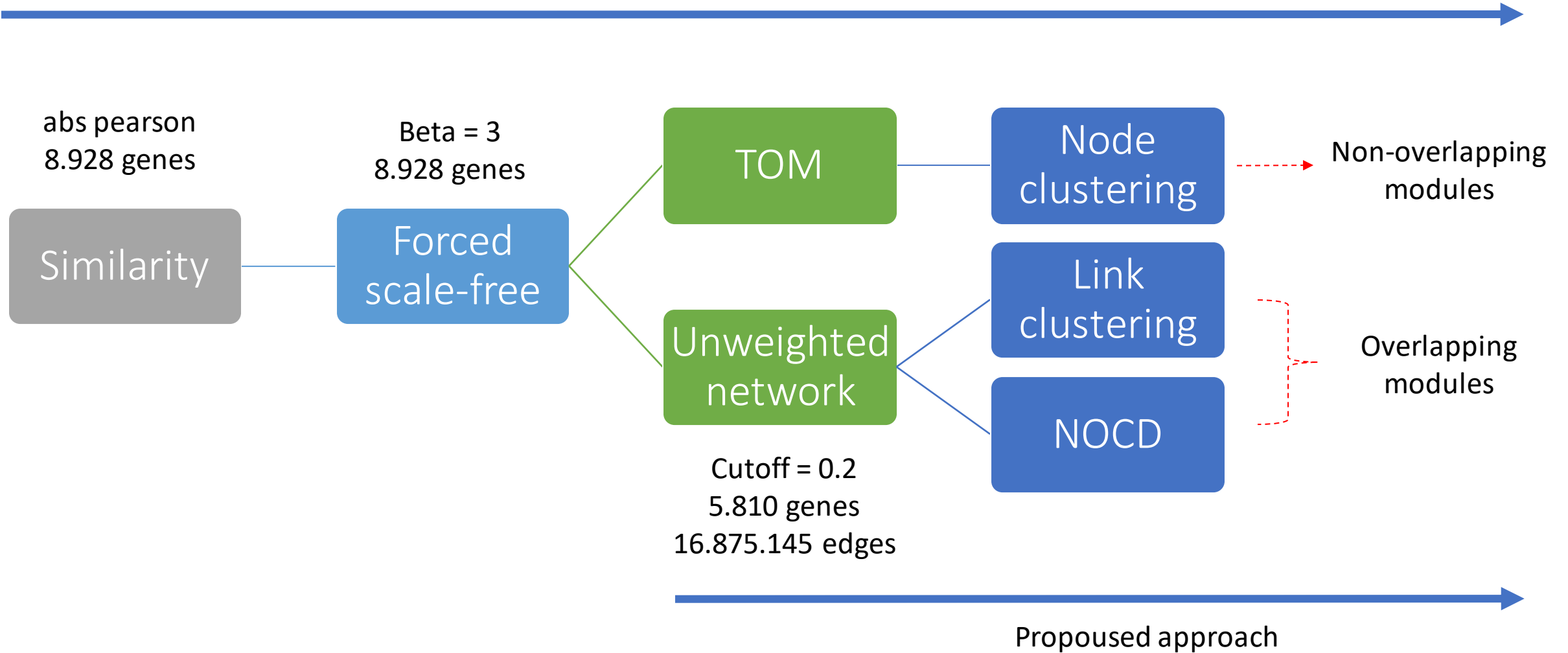
El futuro
es de todos

Gobierno
de Colombia



Modules detection

Usual approach (WGCNA)



In-silico



El futuro
es de todos

Gobierno
de Colombia



Important genes

1PC from LFC matrix
5.143 modules
4.131 genes

Eigengene
per module

Phenotypic trait genes x mod
Shoot K content 3x3
Shoot biomass 3x2
Root biomass 4x1

LASSO



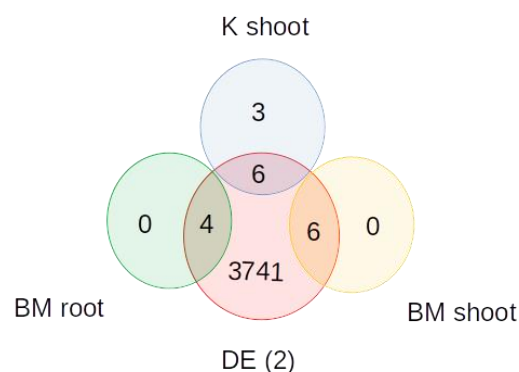
16 DE genes
2 named genes

GO
enrichment

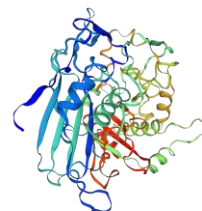


PPI
networks

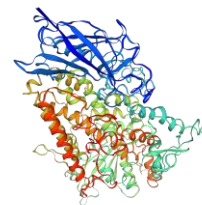
genotype	Module 1	...	Module 5143
101	-1.772555		5.628090
105	1.444241		3.260164
...
4	-1.364746		-1.689424
9	-2.122829		2.862661



SHT2



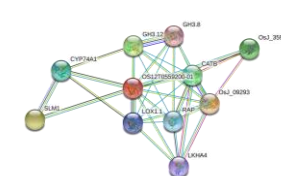
Lipoxygenase



SHT2



Lipoxygenase



In-silico



El futuro
es de todos

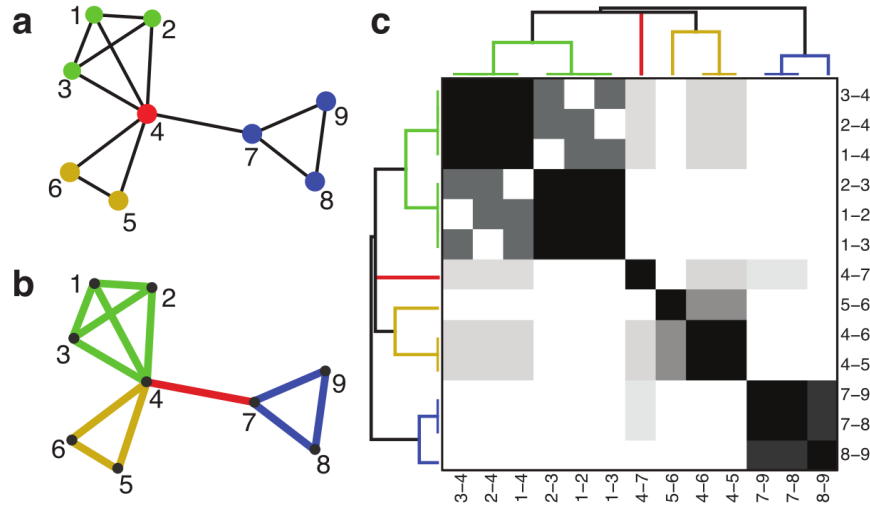
Gobierno
de Colombia



HLC

Input:

Adjacency matrix $A \in \{0, 1\}^{N \times N}$

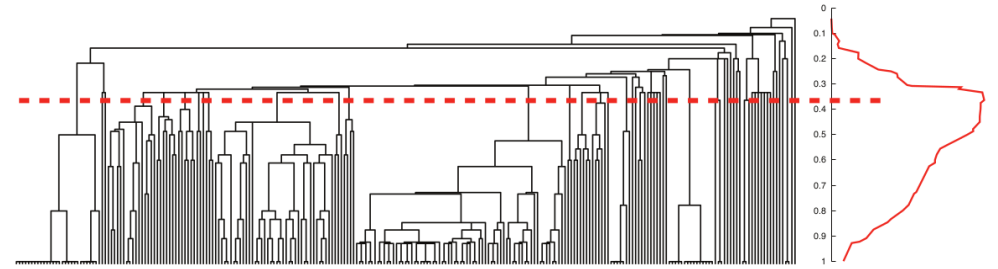


The similarity between two links e_{ik} and e_{jk} is computed using the Jaccard index

$$S(e_{ik}, e_{jk}) = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|}$$

Output:

Affiliation matrix $F \in \{0, 1\}^{N \times C}$



Partition density:

For a network with M links and C subsets

$$D = \frac{2}{M} \sum_c m_c \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}$$

where m_c is the number of links in subset c and n_c is the number of induced nodes.

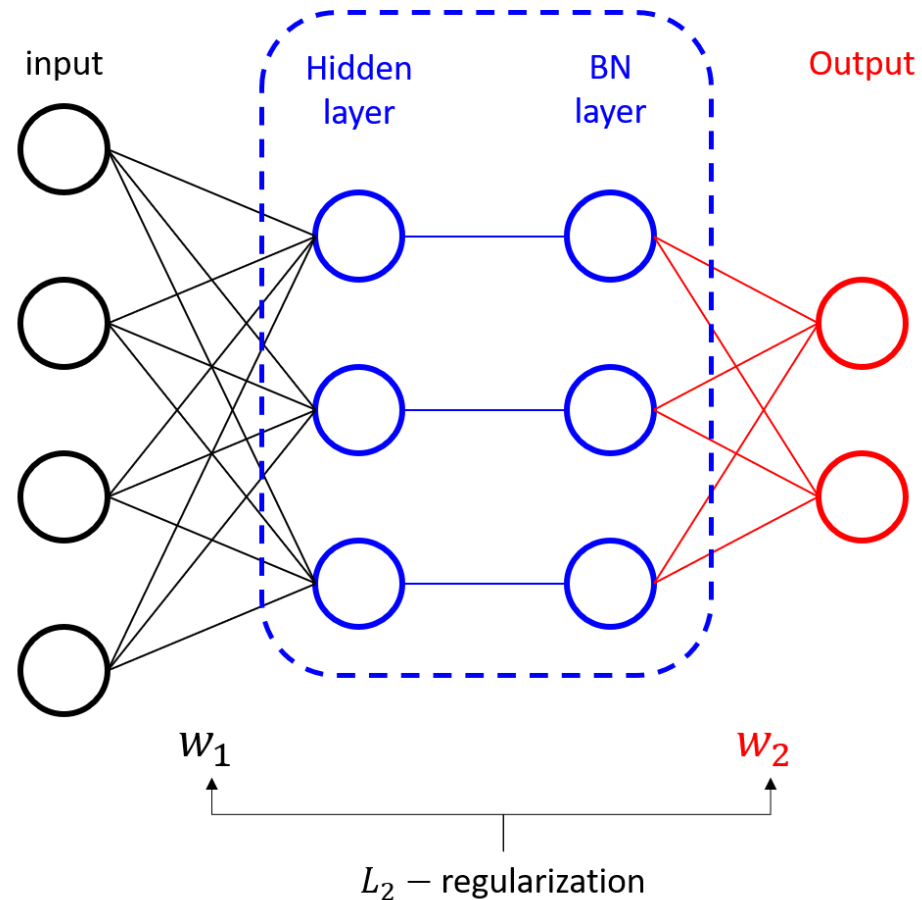
In-silico



El futuro
es de todos

Gobierno
de Colombia





Input:

Adjacency matrix $A \in \{0, 1\}^{N \times N}$

Attribute matrix $X \in \mathbb{R}^{N \times D}$

Real affiliation matrix $Z \in \{0, 1\}^{N \times C}$

Output:

Affiliation matrix $F \in \mathbb{R}_{\geq 0}^{N \times C}$

Threshold $\rightarrow F \in \{0, 1\}^{N \times C}$

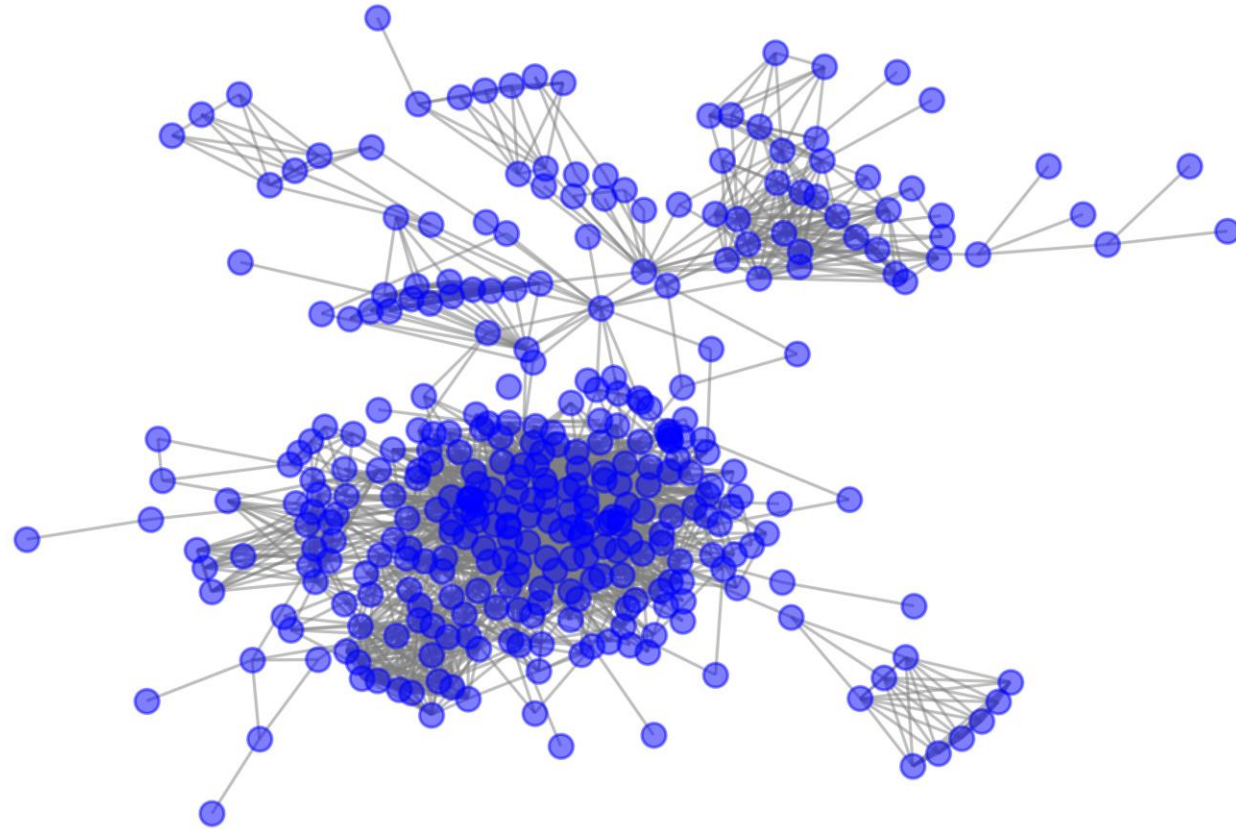
$$F := \text{GCN}_{\theta}(A, X) \\ = \text{ReLU}(\hat{A} \text{ReLU}(\hat{A}XW^{(1)})W^{(2)})$$

N OCD EXAMPLE

Dataset: *fb_0*

$|V| = 347$

$|E| = 2519$



In-silico



El futuro
es de todos

Gobierno
de Colombia



NOCD EXAMPLE

- **Coverage:** what percentage of the edges is explained by at least one community? (i.e. if (u, v) is an edge, both nodes share at least one community) Higher is better.

$$\text{Coverage}(C_1, \dots, C_K) = \frac{1}{|E|} \sum_{u, v \in E} \mathbb{1}[z_u^T z_v > 0]$$

- **Conductance:** average conductance of the detected communities (weighted by community size). Lower is better.

$$\text{outside}(C) = \sum_{u \in C, v \notin C} A_{uv}$$

$$\text{inside}(C) = \sum_{u \in C, v \in C, v \neq u} A_{uv}$$

$$\text{Conductance}(C) = \frac{\text{outside}(C)}{\text{inside}(C) + \text{outside}(C)}$$

$$\text{AvgCond} = \frac{1}{\sum_i |C_i|} \sum_i \text{Conductance}(C_i) \cdot |C_i|$$

- **Density:** average density of the detected communities (weighted by community size). Higher is better.

$$\rho(C) = \frac{\# \text{ existing edges in } C}{\# \text{ of possible edges in } C}$$

$$\text{AvgDensity} = \frac{1}{\sum_i |C_i|} \sum_i \rho(C_i) \cdot |C_i|$$

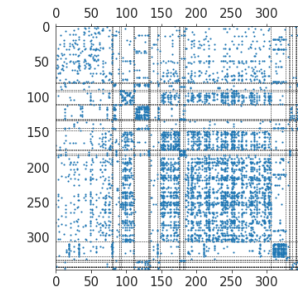
- **Clustering coefficient:** average clustering coefficient of the detected communities (weighted by community size). Higher is better.

$$\text{ClustCoef}(C) = \frac{\# \text{ existing triangles in } C}{\# \text{ of possible triangles in } C}$$

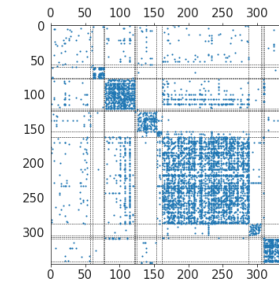
$$\text{AvgCC} = \frac{1}{\sum_i |C_i|} \sum_i \text{ClustCoef}(C_i) \cdot |C_i|$$

- **NMI:** overlapping normalized mutual information. Quantify how similar or different a given set of real clusters are versus the set of clusters found by an algorithm.

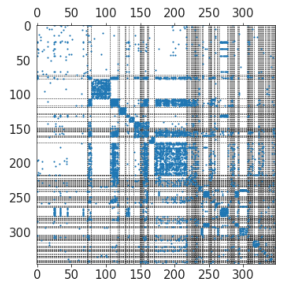
	NOCD	HLC
# Comm.	24	136
Coverage	0.9146	0.8654
Density	2.314e-01	8.395e-01
Conductance	0.2140	0.6877
Clustering	5.947e-02	6.715e-01
NMI	0.1139	0.0473



Real



NOCD



HLC





Aliados



Apoyan



El futuro
es de todos

Gobierno
de Colombia

