

Introducción a Redes Complejas en Biología de Sistemas

Guía 4 – Clustering

- 1) Considere los datos obtenidos en un relevamiento de expresión transcripcional donde a una planta se la sometió a bajas temperaturas durante dos días, con condiciones de luz controladas.

- a. Encuentre la matriz de correlación entre los perfiles transcripcionales y estime a partir de la misma una matriz de similitud S tal que

$$s_{ij} = \frac{1 + \text{cor}(A_i, A_j)}{2}$$

- b. Construya la red de coexpresión génica, utilizando como matriz de adyacencia una versión binarizada de la matriz S , que haga nulo todo elemento de valor menor a 0.95, y 1 todo elemento que supere dicho umbral.
- c. Calcule la partición en clusters de dicha red, obtenida mediante los métodos *infomap* y *fastgreedy* y estime la modularidad de ambas, Visualice ambas redes y compare el tipo y granularidad de las particiones obtenidas.

- 2) Clustering Jerárquico

- a) A partir de los mismos datos de expresión calcule el dendrograma que resulta de realizar un clustering jerárquico con el criterio de *complete linkage*. Visualicelo
- b) Utilizando la red del punto anterior, analice como cambia la modularidad de las particiones obtenidas a partir del dendrograma con diferentes criterios de corte. Puede decir a partir de estos resultados cual sería el número óptimo de clusters que tendría una partición obtenida a partir del dendrograma?

- 3) La comunidad del anillo:

Considere una red unidimensional de N nodos que forma un círculo, donde cada nodo está conectado a sus dos vecinos. Particione el grafo en N_c grupos de tamaño N/N_c .

- a. Calcule la modularidad de la partición obtenida
- b. Cuál es el tamaño de comunidades que maximiza la modularidad?

4) Límite de resolución de la modularidad.

Considere una red conformada por un anillo de n_c M-cliques, cada uno de m nodos y $m(m-1)/2$ enlaces. Cliques vecinos están conectados mediante un único enlace. Esta red tiene una estructura de comunidades natural, donde cada cluster se corresponde con un clique

- Determine la modularidad M_{single} de esta partición natural de la red.
- Determine la modularidad M_{pair} de una partición en la que pares consecutivos de cliques se juntan en un único cluster
- Muestre que sólo si $n_c < 2L$ el criterio de máxima modularidad predice la partición intuitiva, donde

$$L = \frac{n_c m(m-1)}{2} + n_c$$

