



FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

# Estructura comunitaria de la red de neuronas de *C. Elegans*

Sergio Núñez Pazos

2019/2020

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



GRAO DE MATEMÁTICAS

Traballo Fin de Grao

# Estructura comunitaria de la red de neuronas de *C. Elegans*

Sergio Núñez Pazos

Julio 2020

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



# Trabajo propuesto

<b>Área de Coñecemento:</b> Xeometría e Topoloxía
<b>Título:</b> Estructura comunitaria de la red de neuronas de <i>C. Elegans</i>
<b>Breve descripción do contido</b>
El propósito del trabajo es explorar la estructura comunitaria de las redes cerebrales a través del estudio concreto de la red de neuronas del nematodo <i>Caenorhabditis Elegans</i> usando diferentes técnicas y su comparación con diversas redes jerárquicas usadas como redes de referencia.
<b>Recomendacións</b>
<b>Outras observacións</b>



# Índice general

<b>Resumen</b>	<b>VIII</b>
<b>Introducción</b>	<b>XI</b>
<b>1. La red de neuronas de <i>C. Elegans</i></b>	<b>1</b>
1.1. Conectoma de <i>C. Elegans</i> . . . . .	1
1.2. Redes de control . . . . .	3
1.2.1. Internet2 . . . . .	3
1.2.2. Grafo jerárquico de Ravasz . . . . .	3
<b>2. Métodos de detección de la estructura comunitaria</b>	<b>7</b>
2.1. Modularidad y algoritmo de Lovaina . . . . .	7
2.2. Modelos de bloques estocásticos . . . . .	9
2.3. Nodos centrales y nodos conectores . . . . .	12
2.4. Tipos de estructuras . . . . .	14
2.5. Núcleo de una red . . . . .	16
<b>3. Resultados para el conectoma de <i>C. Elegans</i></b>	<b>19</b>
3.1. Estructura comunitaria modular . . . . .	19
3.2. Modelos de bloques estocásticos . . . . .	27
3.3. Centro y núcleo . . . . .	30
<b>4. Resultados para las redes de control</b>	<b>37</b>
4.1. Internet2 . . . . .	37
4.1.1. Algoritmo de Lovaina . . . . .	37
4.1.2. Modelos de bloques estocásticos . . . . .	43
4.1.3. Centro y núcleo . . . . .	49
4.2. Grafo jerárquico de Ravasz . . . . .	53
4.2.1. Algoritmo de Lovaina . . . . .	53

4.2.2. Modelos de bloques estocásticos . . . . .	55
4.2.3. Centro y núcleo . . . . .	58
<b>5. Comparación de métodos</b>	<b>63</b>
5.1. Algoritmo de Lovaina . . . . .	63
5.2. Modelos de bloques estocásticos . . . . .	65
5.3. Tipo de estructura y dependencia del algoritmo . . . . .	66
<b>6. Conclusiones</b>	<b>73</b>
<b>Bibliografía</b>	<b>75</b>





## Resumen

Este trabajo tiene tres objetivos principales. En primer lugar estudiamos la estructura comunitaria del conectoma del nematodo *C. elegans* por medio de dos técnicas diferentes: el algoritmo de Lovaina y los modelos de bloques estocásticos. Además, estudiamos si existe alguna relación entre la distribución modular obtenida y la partición del conectoma en ganglios y grupos funcionales. En segundo lugar, mediante la implementación de estos algoritmos de detección de la estructura comunitaria sobre redes de control, cuya distribución modular es conocida, verificamos en qué casos los métodos empleados son más eficaces. Por último, tratamos de definir sobre cada red un *núcleo*, es decir, un conjunto de nodos encargados de su correcto funcionamiento, de manera que si este desaparece de la red, esta no podría realizar las funciones para las que fue diseñada.

## Abstract

This work has three different objectives. First of all, we study the *C. elegans* connectome and, using different techniques like the Louvain algorithm and stochastic block modeling technics (SBM), we try to find his community structure. Moreover, we study if there is a relationship between the obtained modular distribution and the partitioning of the connectome in ganglia and functional groups. Secondly, applying these community detection algorithms to benchmark networks as Internet2 and Ravasz's hierarchical graph for which we know it's community structure, we check in which cases each algorithm works better. Finally, we try to define for each network a *core*, that is, we try to set a collection of nodes that are responsible for each network to make it work properly. Moreover, if we remove this subset from the network, it should stop working.



# Introducción

A lo largo de este trabajo tratamos de entender la estructura de la red neuronal del nematodo *Caenorhabditis elegans*, al mismo tiempo que tratamos de abstraer conclusiones válidas para los conectomas de otras especies. Para ello aplicamos distintos métodos para detectar la estructura comunitaria y estudiar la arquitectura de la red. En primer lugar empleamos el *algoritmo de Lovaina*, un método basado en el concepto de *Q-modularidad* introducido por Newman y Girvan y donde el algoritmo busca una estructura comunitaria que maximice ese valor. Al hacerlo, proporciona una división del conectoma en módulos que maximizan la densidad de conexiones internas y minimizan las sinapsis entre neuronas de distintos módulos. Por otra parte, también empleamos *modelos de bloques estocásticos*, un nuevo método que trata de identificar la estructura comunitaria de las redes en base a los patrones de conectividad de los nodos de la red.

La idea de modularidad presentada por Newman y Girvan y la aplicación del algoritmo de Lovaina ha inducido a pensar que las redes neuronales tienen una estructura *asortativa*, es decir, con alta densidad de conexiones en las comunidades de la red y una baja densidad de conexiones entre los nodos de distintos módulos de la red. Sin embargo, la aparición de los modelos de bloques estocásticos está cambiando esta idea proponiendo una estructura *disasortativa* derivada de la ejecución de funciones de forma coordinada en la red. De todos modos, nuestros resultados ponen en cuestión ambas ideas pues el tipo de estructura comunitaria de una red depende del algoritmo usado.

Una vez estudiada la estructura comunitaria del conectoma de *C. elegans* mediante ambos algoritmos, intentamos relacionar los resultados obtenidos con las divisiones anatómicas, en ganglios, y funcionales de las neuronas del conectoma. Tiene especial interés identificar tanto la subred por la que transita la mayoría de la información (*hourglass weist*) como el conjunto de neuronas que concentran la mayoría de las sinapsis (*rich club*). No obstante, nos interesa definir un conjunto de neuronas, al que llamaremos *núcleo*, que esté caracterizado por el hecho de que la red dejaría de funcionar si se elimina.

En la literatura estos métodos se aplican únicamente a la red de *C. elegans* o, a lo sumo, se hace una comparativa con otras redes cerebrales de otras especies. Sin embargo, a

menudo estas no son redes neuronales al completo, sino conectomas a *media escala*, donde los nodos no son neuronas, sino regiones cerebrales de interés (ROIs) y las conexiones que se establecen al realizar determinadas tareas. Por ello, nuestro tercer objetivo es testar la eficiencia real de los algoritmos usados en el estudio de redes muy complejas. Para ello empleamos dos redes de control, una tecnológica y otra artificial, con una estructura topológica y geométrica conocida con las que comprobar hasta qué punto estos métodos son capaces de identificar la estructura y la arquitectura de nuestras redes.

# Capítulo 1

## La red de neuronas de *C. Elegans*

### 1.1. Conectoma de *C. Elegans*

*Caenorhabditis elegans* (*C. elegans* en lo sucesivo) es una especie de gusano transparente de aproximadamente 1 milímetro de longitud que forma parte de la familia de los nematodos. Su importancia en la investigación biológica es mayúscula pues se emplea en distintos campos, desde la investigación en biología molecular hasta el estudio farmacológico. Parte de su interés científico se debe a la sencillez de su red cerebral, pues el conectoma del ejemplar hermafrodita apenas cuenta con 302 neuronas y 3465 sinapsis, lo que permite un estudio minucioso y detallado de la red. En la Figura 1.1 mostramos una representación del conectoma de *C. elegans* tomada de OpenWorm [12].

El conectoma del *C. elegans* adulto fue descrito por primera vez por J. G. White et al. en [21], donde se decía que la red neuronal constaba de 282 neuronas somáticas. Sin embargo, actualmente se sabe que la red neuronal consta de 302 neuronas. Esta nueva distribución fue descrita por L. R. Varshney et al. en [20] y cuenta con una componente conexa de 300 neuronas y dos neurocitos aislados (CANL y CANR). Nuestro estudio se centra en el cálculo de la estructura comunitaria de la componente conexa con 300 neuronas, por lo que en lo sucesivo nos referiremos a esta como si del conectoma al completo se tratase. Además, en dicha descripción se establecen dos criterios de clasificación de las neuronas de *C. elegans*. La primera se basa en una distribución anatómica de las neuronas de la red en ganglios, que son los conjuntos posicionales en los que las neuronas se agrupan a lo largo del conectoma. Se distinguen 11 ganglios distintos: faringe, anterior, dorsal, ventral, lateral, retrovesicular, cordón nervioso ventral o ventral central, posterolateral, preanal, dorsorectal y lumbar. Especial interés tiene la disposición de los ganglios en *C. elegans*, que se encuentran repartidos a lo largo de todo el cuerpo del mismo, haciendo posible una

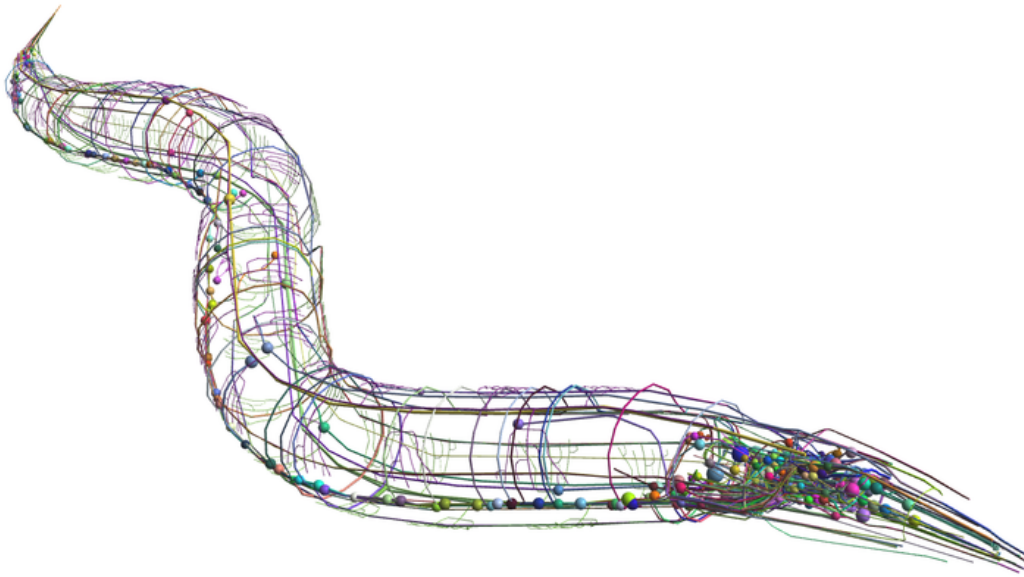


Figura 1.1: Representación del conectoma de *C. elegans*.

representación de la posición que estos ocupan en el cuerpo del gusano. En la Figura 1.2 podemos observar esta disposición. Por otra parte, también se define una clasificación funcional de las neuronas del conectoma. Esta se basa en la función que realiza cada neurocito en la red. En ella se distinguen tres tipos de neuronas: *sensoriales*, encargadas de interceptar estímulos externos e introducirlos en el conectoma, *interneuronas*, encargadas de transmitir los impulsos nerviosos dados por las neuronas sensoriales a lo largo del conectoma, y *motoras*, que coordinan su movimiento en base a los estímulos recibidos.

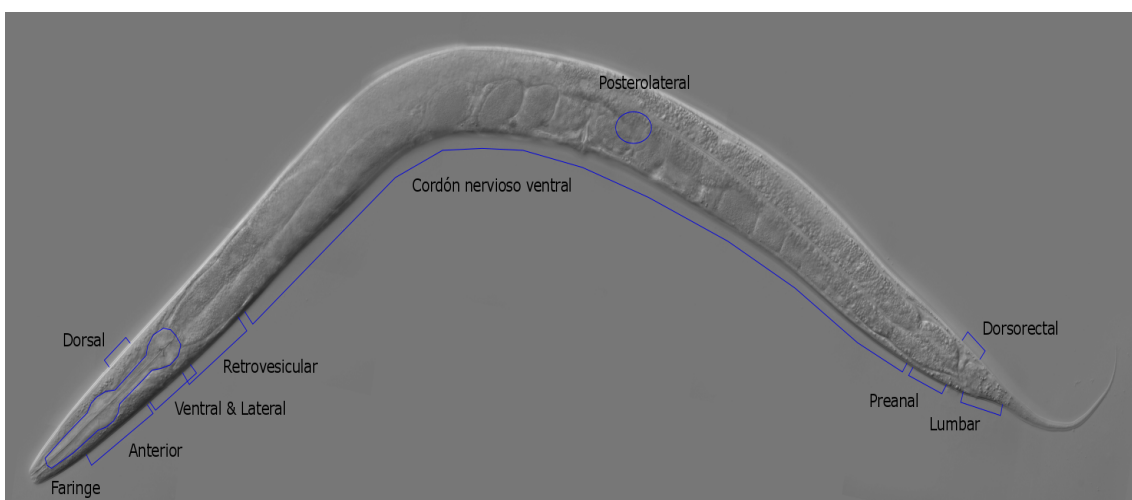


Figura 1.2: Representación de los ganglios de *C. elegans*.

Al aplicar los métodos para el cálculo de la estructura comunitaria comparamos la partición obtenida con la división anatómica y funcional antes expuesta, tratando de buscar una relación comunitaria de los grupos de neuronas.

## 1.2. Redes de control

### 1.2.1. Internet2

Por otra parte, aplicamos los mismos métodos a redes no biológicas que cuentan con una estructura topológica conocida y que nos sirven como redes de control de los diferentes algoritmos de detección de estructuras comunitarias. Internet2 se fundó en Estados Unidos en 1996 y, mediante este consorcio, universidades de todo el país colaboran en la investigación y la educación. Según la descripción de [1], referida a la situación en abril de 2013, en esta organización sin ánimo de lucro participan 311 universidades con 320 conexiones de fibra óptica. Cabe destacar que, aunque al comienzo el proyecto se limitaba a universidades estadounidenses, actualmente busca la internacionalización a través de convenios con otros países.

En lo que a la red se refiere, procede señalar su peculiar distribución. Esta se divide en dos conjuntos: un bloque central formado por 60 universidades, que conforma un núcleo geográfico dispuesto en 6 ciclos por el que se transporta la información a lo largo de la red, y una periferia de 351 instituciones académicas que actúan como nodos terminales que se conectan a la estructura cíclica central. Esta distribución caracteriza a las redes de tipo *centro-periferia*, que cuentan con un núcleo más o menos interconectado al que se unen nodos terminales mediante enlaces simples. En la Figura 1.3 presentamos una representación del bloque central de la red sacada de [1].

### 1.2.2. Grafo jerárquico de Ravasz

Por último consideramos una segunda red de control asociada a un grafo jerárquico. Esta red fue propuesta por E. Ravasz y sus coautores en [15] y se basa en una idea de construcción de redes de forma inductiva, dando lugar a una estructura jerárquica según la posición ocupada por los nodos en cada iteración. Hay una estructura central que actúa como base y que se replica sucesivamente dando lugar a una periferia que aumenta de complejidad a medida que se itera el proceso. Esta red tiene escaso valor práctico pues a diferencia de las otras dos, no representa la red asociada a un sistema real. Sin embargo, a nivel teórico su estructura jerárquica nos permite contrastar el correcto funcionamiento de los algoritmos de detección de la estructura comunitaria.

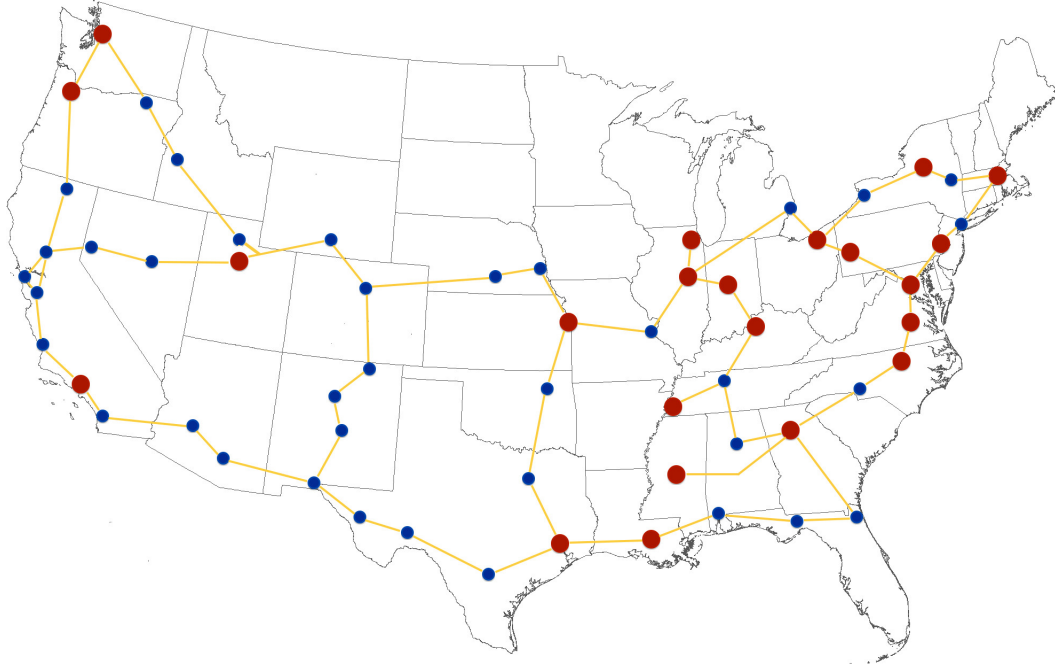
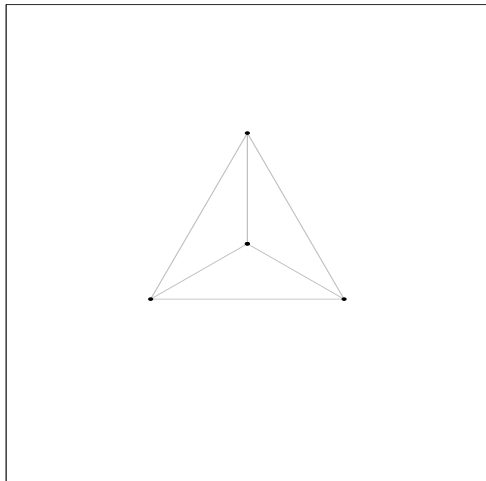


Figura 1.3: Representación de la red tecnológica Internet2.

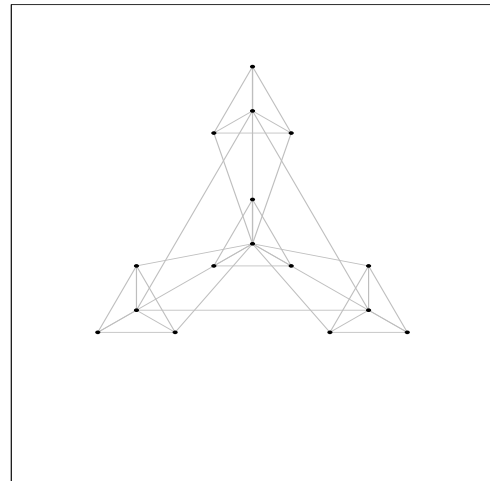
A continuación presentamos detalladamente la estructura de la red de Ravasz y junto a esta la representación del grafo hasta orden 4 en la Figura 1.4.

El grafo de orden 1 consta de 4 nodos: el central y tres periféricos colocados formando un tetraedro. El nodo central está conectado con los tres nodos periféricos, mientras que estos están conectados entre sí. El grafo de orden 2 está formado por 4 copias del de orden 1, una central y otras tres alrededor, consiguiendo de nuevo una estructura de tipo centro-periferia con un bloque central y tres bloques que conforman la periferia. En ella el nodo central del grafo central está conectado con la periferia de la periferia, mientras que los centros de los grafos periféricos están conectados entre sí. Repitiendo el proceso inductivamente obtenemos los grafos jerárquicos con órdenes crecientes.

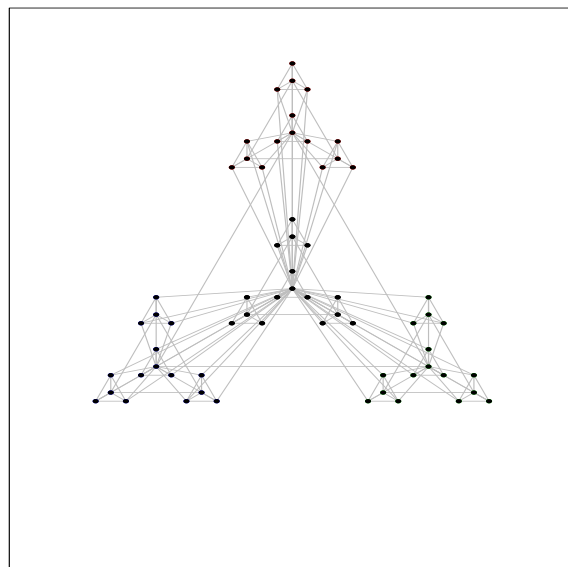
De esa forma, en base a la topología de la red, podemos dividir el grafo jerárquico según dos criterios. En primer lugar, mediante la división natural en cuatro bloques, donde uno corresponde al bloque central y los otros tres a los bloques periféricos, siendo estos idénticos en cuanto a la posición relativa de sus nodos. Por otra parte, podemos definir una división jerárquica del grafo, según la posición ocupada por los nodos que lo conforman.



(a) Grafo jerárquico de orden 1.

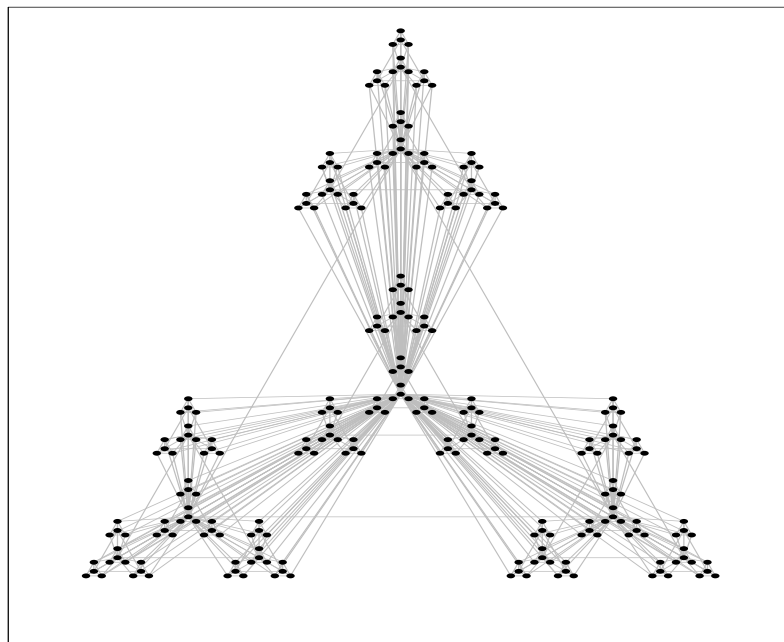


(b) Grafo jerárquico de orden 2.



(c) Grafo jerárquico de orden 3.

Figura 1.4: Representación de los grafos jerárquicos de los tres primeros órdenes.



(d) Grafo jerárquico de orden 4.

Figura 1.4: Representación del grafo jerárquico de orden 4.

## Capítulo 2

# Métodos de detección de la estructura comunitaria

### 2.1. Modularidad y algoritmo de Lovaina

**Definición 2.1.** Una *estructura comunitaria* de una red es una partición de la misma en conjuntos de nodos en los que se maximiza la densidad de conexiones dentro de los módulos respecto al número de enlaces con otros nodos de la red. Además, los conjuntos dados por esta partición también guardan una relación en su conectividad, pues dos nodos de un mismo grupo tienen más posibilidades de estar conectados que nodos de distintos grupos. A los conjuntos obtenidos en esta partición se les denominan *comunidades* o *módulos*.

Para llevar a cabo el cálculo de la estructura comunitaria del nematodo *C. elegans* comenzamos definiendo el concepto de modularidad. Asociada a cualquier partición tenemos la noción de *Q-modularidad* introducida por Newman y Girvan en [11]. Esta cantidad es la diferencia entre el número de conexiones presentes dentro de los módulos y la cantidad de conexiones que serían esperables en una red equivalente con las aristas distribuidas aleatoriamente. Intuitivamente es una medida de la bondad de ajuste de la red a una división comunitaria.

Más precisamente, tenemos la siguiente definición:

**Definición 2.2.** Si  $G = (V, E)$  es un grafo con  $V = \{1, \dots, N\}$ ,  $E$  el conjunto de aristas del grafo y  $\mathcal{P} = \{P_k\}$  una partición del conjunto de vértices  $V$ , se define la *Q-modularidad* de una partición  $\mathcal{P}$  como:

$$Q = \frac{1}{2|E|} \sum_{i,j \in P_k} \left[ A_{ij} - \frac{k_i k_j}{2|E|} \right],$$

donde  $A_{ij}$  denota la entrada en la posición  $i, j$  de la matriz de adyacencia del grafo,  $k_i$  denota el grado del vértice  $i$ ,  $|E|$  es el número de aristas y  $\frac{k_i k_j}{2|E|}$  es la probabilidad de conexión entre el vértice  $i$  y el  $j$ . El factor  $\frac{1}{2|E|}$  se emplea para normalizar el valor obtenido.

Claramente  $Q \in [-\frac{1}{2}, 1]$ , donde una modularidad alta se corresponde con módulos  $P_k$  densamente conectados, mientras que un valor bajo implica una baja concentración de conexiones entre los vértices de los módulos  $P_k$ .

Una partición  $P_k$  que maximice el valor de  $Q$  se suele considerar como modelo de la estructura comunitaria del grafo, agrupando los nodos de tal forma que se maximice la densidad de las conexiones en cada grupo y se minimicen los enlaces entre nodos de distintos grupos. Se llama *Q-modularidad* de una red a ese valor máximo.

El *algoritmo de Lovaina* se basa en esta idea para proponer un modelo de estructura comunitaria de una red. Este algoritmo consta de las siguientes etapas:

1. Inicialmente tomamos cada nodo como un módulo independiente. A partir de esto, para cada uno de ellos, calculamos la variación de modularidad que supondría agrupar cada vértice en una comunidad distinta a la suya. Usando esto, agrupamos cada nodo en aquella comunidad que reporte un mayor incremento de modularidad o, si se diera el caso y no aumentase este valor, el nodo permanecería en su módulo inicial.
2. En la siguiente etapa, construimos una nueva red en la que cada módulo pasa a representar un nodo y las conexiones internas entre nodos de un módulo lazos. A esta nueva estructura podemos aplicarle el proceso descrito en la primera etapa.
3. Repitiendo el proceso iterativamente podemos dar por finalizado el algoritmo si al realizar una iteración no se encuentra ninguna posibilidad de aumentar la modularidad agrupando las comunidades o si tan solo queda un nodo en la red.

Así, procediendo según el esquema anterior, obtenemos un modelo para la estructura comunitaria dado por el algoritmo de Lovaina sobre las redes estudiadas.

Cabe destacar que este algoritmo no es determinístico, sino que en diversas iteraciones puede dar resultados diferentes. Por ejemplo, puede haber posibles variaciones en los módulos obtenidos en función del orden en el que se tomen los nodos en el primer paso.

Una característica de este algoritmo es que se lleva a cabo mediante un procedimiento jerárquico, donde en cada iteración vamos obteniendo módulos formados por los que fueron obtenidos en los pasos anteriores.

Conviene señalar que el procedimiento se puede simplificar, pues a la hora de ejecutar el primer paso, no es necesario comprobar la variación de modularidad que supondría la

inclusión de un nodo a una comunidad que no le sea adyacente, pues considerar módulos desconexos no maximiza la modularidad de la red (Proposición 2.3). Por lo tanto, el algoritmo de Lovaina siempre proporciona módulos conexos.

**Proposición 2.3.** *Los módulos dados por el algoritmo de Lovaina son conexos.*

*Demostración.* Tomemos dos módulos desconexos y veamos que agruparlos no supone un beneficio a nivel de modularidad y, por tanto, no conforman un módulo desconexo.

Llamemos  $Q_{m,n}$  a la modularidad tomando los módulos  $m$  y  $n$  por separado y  $Q_{mn}$  a la modularidad obtenida agrupando los módulos. Como las comunidades son desconexas se verifica que  $A_{ij} = 0$  para cada  $i \in P_m$  y  $j \in P_n$ . Además denotamos por  $Q$  la modularidad de los restantes módulos. Luego:

$$\begin{aligned}
Q_{m,n} &= \frac{1}{2|E|} \left( \sum_{i,j \in P_m} [A_{ij} - \frac{k_i k_j}{2|E|}] + \sum_{i,j \in P_n} [A_{ij} - \frac{k_i k_j}{2|E|}] \right) + Q \\
&> \frac{1}{2|E|} \left( \sum_{i,j \in P_m} [A_{ij} - \frac{k_i k_j}{2|E|}] + \sum_{i,j \in P_n} [A_{ij} - \frac{k_i k_j}{2|E|}] \right) + \frac{1}{2|E|} \sum_{i \in P_m, j \in P_n} [-\frac{k_i k_j}{2|E|}] + Q \\
&= \frac{1}{2|E|} \left( \sum_{i,j \in P_m} [A_{ij} - \frac{k_i k_j}{2|E|}] + \sum_{i,j \in P_n} [A_{ij} - \frac{k_i k_j}{2|E|}] + \sum_{i \in P_m, j \in P_n} [A_{ij} - \frac{k_i k_j}{2|E|}] \right) + Q = Q_{mn}
\end{aligned} \tag{2.1}$$

Es decir, agrupar dos módulos no adyacentes no supone un aumento de la Q-modularidad. Por tanto, el algoritmo de Lovaina nunca devuelve una comunidad desconexa.  $\square$

De esta forma, hemos comprobado que el algoritmo de Lovaina trata de detectar la estructura comunitaria de una red a partir del concepto de modularidad, obteniendo una partición en comunidades que agrupan nodos con una alta densidad de conexiones intracomunitaria. Una de las cuestiones que nos planteamos es saber hasta qué punto este procedimiento tiene en cuenta las propiedades espaciales de la red, tanto topológicas como geométricas, o ciertos patrones usados en su construcción. Como contraparte de este procedimiento presentamos otro método de detección de la estructura comunitaria, que no se basa en el concepto de la Q-modularidad, sino que agrupa los nodos siguiendo patrones de conectividad similares.

## 2.2. Modelos de bloques estocásticos

Los *modelos de bloques estocásticos* fueron introducidos en los años 1980 en el campo de los estudios sociológicos, pero ahora son frecuentemente empleados en el análisis de redes, combinados con técnicas de *machine learning*, sistemas complejos y física estadística.

Convencionalmente, los SBM están definidos para grafos simples no orientados, aunque pueden ser generalizados para grafos orientados con bucles. Además, se han desarrollado versiones en las que las aristas cuentan con pesos.

Comenzamos definiendo estos modelos. Un SBM básico está definido por una terna  $(k, z, M)$ , donde :

- $k$  es un escalar que denota el número de módulos dados por la estructura comunitaria de la red.
- $z$  es un vector de orden  $N = |V|$  donde  $z_i \in \{1, 2, \dots, k\}$  denota la comunidad a la que pertenece el vértice  $i$ .
- $M$  denota una matriz  $k \times k$  donde cada entrada  $M_{u,v}$ , con  $u, v \in \{1, \dots, k\}$  es un *bloque estocástico* que representa la probabilidad de que un nodo del módulo  $u$  se conecte con un nodo del módulo  $v$ .

*Observación 2.4.* Por definición, los nodos de cualquier módulo son estocásticamente equivalentes, pues nodos de una misma comunidad siguen los mismos patrones de conectividad. De modo más preciso, cada módulo es un *grafo aleatorio* de tipo Erdős-Rényi donde la probabilidad de que dos nodos estén conectados es precisamente  $M_{z_i, z_j}$ . Además, como los valores de la matriz  $M$  determinan la densidad de conexiones entre grupos, la estructura entre módulos corresponde a un grafo bipartito aleatorio [5].

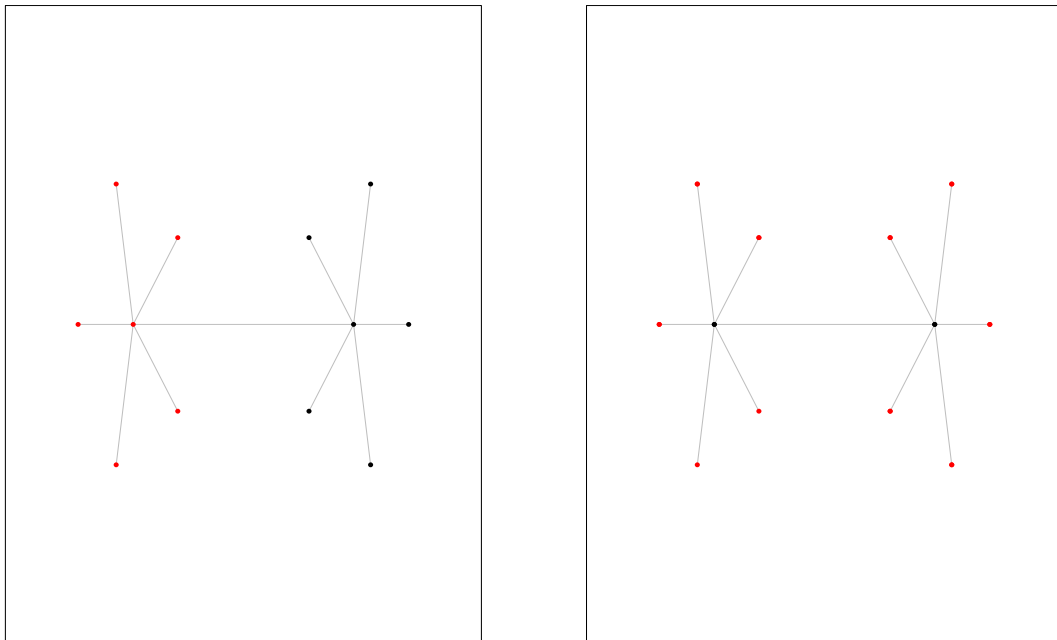
Una vez definida esta terna podemos generar una red en función de los valores asignados a  $k, z$  y  $M$ . Esto se hace tomando dos vértices  $i$  y  $j$  de la red y conectándolos o no con una probabilidad  $M_{z(i), z(j)}$ . De esta forma, dado  $G = (V, E)$ , podemos inferir en sentido estadístico su estructura comunitaria con SBM. Así, fijado  $k$  tenemos la función de verosimilitud

$$\begin{aligned}
 \mathcal{L}(G|M, z) &= \prod_{i,j} \mathbb{P}(i \rightarrow j|M, z) \\
 &= \prod_{(i,j) \in E} \mathbb{P}(i \rightarrow j|M, z) \prod_{(i,j) \notin E} 1 - \mathbb{P}(i \rightarrow j|M, z) \\
 &= \prod_{(i,j) \in E} M_{z_i, z_j} \prod_{(i,j) \notin E} (1 - M_{z_i, z_j})
 \end{aligned} \tag{2.2}$$

basta entonces procurar que  $M$  y  $z$  maximicen esta cantidad.

De esa forma, mediante un método estadístico basado en la inferencia, se buscan los valores del vector  $z$  y la matriz  $M$  que maximizan la verosimilitud de modo que el modelo se ajusta mejor al grafo original  $G$ .

Como dijimos, los SBM no buscan la partición en grupos con alta densidad de conexiones internas acorde al concepto de  $Q$ -modularidad, como hacía el algoritmo de Lovaina, sino que busca un reparto en grupos donde los vértices sigan patrones de conexión similares desde un punto de vista estadístico. Así, por ejemplo considerando dos grafos estrella conectados por una arista que va entre los nodos centrales de cada una, obtenemos que mientras el algoritmo de Lovaina separa ambos grafos en dos comunidades, si a los modelos de bloques estocásticos le pidiesemos que nos diera una estructura comunitaria con 2 módulos, esta separaría los dos nodos centrales de los periféricos, diferenciando los patrones de conectividad de cada vértice. Esta distribución se puede ver en la Figura 2.1.



(a) Estructura comunitaria según el algoritmo de Lovaina.

(b) Estructura comunitaria según SBM.

Figura 2.1: Estructura comunitaria del grafo estrella.

De esa forma, podemos observar que, a diferencia del algoritmo de Lovaina, aplicando SBM podemos obtener estructuras comunitarias cuyos módulos son desconexos, pues SBM tan solo busca grafos aleatorios que en función de los valores de  $z$  y de  $M$  se ajusten a la red real y no grupos de alta densidad de conexiones, por lo que los módulos pueden llegar a salir desconexos.

### 2.3. Nodos centrales y nodos conectores

Por otra parte, una vez introducidos los métodos empleados para el cálculo de la estructura comunitaria de una red, nos interesamos por el estudio individual de los nodos que la conforman. Para ello presentamos diversos conceptos que nos permiten clasificar los nodos de una red en base a sus patrones de conectividad. Comenzamos definiendo el coeficiente de participación introducido por R. Guimerà y L. A. N. Amaral en [8].

**Definición 2.5.** Sea  $\mathcal{P}$  una partición del conjunto de nodos  $V = \{1, \dots, N\}$  de una red obtenida por cualquier algoritmo de detección de la estructura comunitaria donde  $m$  es el número de módulos,  $d_i$  el grado del nodo  $i$  y  $d_{ic}$  el número de conexiones del vértice  $i$  con nodos del módulo  $c$ . Se llama *coeficiente de participación* de un nodo  $i$  a la proporción:

$$\mathcal{P}_i = 1 - \sum_{c=1}^m \left( \frac{d_{ic}}{d_i} \right)^2.$$

Podemos observar que cuantas más conexiones tenga el nodo  $i$  con otros módulos, mayor es el coeficiente de participación, mientras que este disminuye si las conexiones son mayoritariamente entre nodos de su comunidad. Originalmente esta cantidad se introdujo para la partición proporcionada por el algoritmo de Lovaina, sin embargo esta fue extendida a las particiones dadas por otros algoritmos de detección de la estructura comunitaria.

**Ejemplo 2.6.** Tomemos  $V_1$  un nodo terminal asociado a una red arbitraria. Claramente tenemos que  $d_1 = 1$ , por ser terminal, y  $d_{1c} = 0$  para todo  $c$  salvo para un  $c_0$  tal que  $d_{1c_0} = 1$ . De esa forma,  $\sum_{c=1}^m \left( \frac{d_{1c}}{d_1} \right)^2 = \left( \frac{d_{1c_0}}{d_1} \right)^2 = 1$  y, por tanto,  $\mathcal{P}_1 = 0$ .

**Ejemplo 2.7.** Tomemos el grafo representado en la Figura 2.1 y la estructura comunitaria dada por el algoritmo de Lovaina. Denotamos por NT a los diez nodos terminales de la red y por NC a los dos nodos centrales. Calculemos el coeficiente de participación de estos.

Como NT son nodos terminales, por el Ejemplo 2.6 tenemos que  $\mathcal{P}_i = 0$  para todo  $i \in NT$  de la red.

Por otra parte, por simetría tenemos que  $\mathcal{P}_{NC_1} = \mathcal{P}_{NC_2}$ . De esa forma  $\mathcal{P}_{NC_1} = \mathcal{P}_{NC_2} = 1 - \sum_{c=1}^2 \left( \frac{d_{ic}}{d_i} \right)^2 = 1 - \left( \frac{5}{6} \right)^2 - \left( \frac{1}{6} \right)^2 = \frac{5}{18}$ .

En [18] O. Sporns et al. llaman *conectores* a los nodos que tienen una participación elevada y *provinciales* a los que cuentan con un coeficiente menor. En particular fijan el siguiente criterio: un nodo es conector si  $P_i \geq 0,3$  y provincial si  $P_i < 0,3$ . Las neuronas conectoras son las encargadas de transmitir la información de su módulo a todo el conectoma, mientras las provinciales transmiten la información a nivel local.

*Observación 2.8.* A la vista del Ejemplo 2.6 podemos afirmar que todos los nodos terminales son provinciales, pues su coeficiente de participación es nulo.

Dado un grafo, la distancia entre nodos es el menor número de aristas que separan dichos vértices. Esta idea permite definir una métrica sobre los grafos no orientados, extendiendo su definición a lo largo del grafo. Además, conviene recordar que un *camino geodésico* entre dos nodos de un grafo es aquel que une los vértices mediante el menor número de aristas posibles.

Este concepto permite definir el *coeficiente de centralidad*, introducido por L. Freeman en 1977.

**Definición 2.9.** El *coeficiente de centralidad* de un nodo se define como la proporción de caminos geodésicos de la red que pasan por él. Es decir, el coeficiente de centralidad mide el promedio de veces que un nodo actúa como puente en un camino geodésico entre dos nodos arbitrarios de la red. Dado el conjunto de nodos  $V = \{1, \dots, N\}$ , la centralidad de  $i$  es:

$$C_i = \frac{2}{(N-1)(N-2)} \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \quad i \in V$$

donde  $\sigma_{jk}(i)$  denota el número de caminos geodésicos de  $j$  a  $k$  que pasan por el nodo  $i$  y  $\sigma_{jk}$  representa el número total de caminos geodésicos de  $j$  a  $k$ . El coeficiente  $\frac{2}{(N-1)(N-2)}$  se emplea para normalizar el valor obtenido, de forma que  $C_i \in [0, 1]$ .

*Observación 2.10.* Los nodos terminales tienen coeficiente de centralidad nulo, pues ningún camino geodésico del que no sean extremo pasa por ellos.

En [18] se llaman *nodos centrales* a aquellos que tienen un coeficiente de centralidad  $C_i \geq 0,1$ . Los nodos centrales forman lo que llamamos *centro de la red*. Estos nodos son los encargados de transmitir la información a lo largo de la red con la mayor celeridad posible. Sin embargo, en nuestro caso, modificamos el criterio de [18] considerando centrales aquellos nodos que superen el valor medio de la centralidad de la red. Esta nueva asignación no cambia la interpretación de estos nodos, sino que tan solo implica un reajuste de este concepto a las redes estudiadas, puesto que las redes más complejas cuentan con multitud de conexiones, lo que provoca una reducción considerable del coeficiente de centralidad de los nodos. Conviene resaltar que el criterio fijado en [18] se aplica a un conectoma, precisamente el conectoma humano, pero a media escala. Esto significa que los nodos no representan neuronas, sino regiones de la corteza cerebral con una área media de  $1,5 \text{ cm}^2$  y un tamaño de la red aproximado de  $N = 998$ .

Por otra parte, usamos los caminos geodésicos para introducir una subred encargada de conectar los distintos módulos y de transmitir la información entre ellos.

**Definición 2.11.** Llamamos *red de conectores* a la componente conexa de tamaño máximo de la subred formada por la unión de todos los caminos geodésicos que conectan dos nodos conectores arbitrarios.

La importancia de esta subred es capital, pues sin ella no sería posible transmitir la información entre los distintos módulos o al menos entre una parte sustancial de esos módulos.

## 2.4. Tipos de estructuras

En este apartado estudiamos la arquitectura y la topología de las distintas redes, basándonos en sus patrones de conectividad. En la literatura, se suelen distinguir tres tipos de estructuras:

1. *Asortativa.* Una red es *asortativa* si sus comunidades forman conjuntos densamente conectados internamente y con una baja cantidad de conexiones externas. Según algunos autores, como en [4], las comunidades con este tipo de estructuras se caracterizan por ser robustas ante perturbaciones y adaptarse fácilmente a cambios evolutivos. Además, por su disposición espacial, los módulos de estas redes están especializados en la realización de funciones que luego combinan entre sí para el funcionamiento completo de la red. Esta idea es cuestionada en el Capítulo 5 al estudiar la dependencia de la estructura comunitaria del algoritmo utilizado.

El algoritmo de Lovaina busca precisamente encontrar una estructura *asortativa*, pues al maximizar la  $Q$ -modularidad los módulos encontrados siguen el patrón de núcleos con una alta densidad de conexiones.

2. *Núcleo-Periferia.* Una red es de tipo *núcleo-periferia* cuando esta cuenta con un conjunto de nodos con una alta densidad de conexiones que actúa como un núcleo al que luego se conecta un conjunto de vértices de bajo *grado* que actúan como periferia. Esta estructura permite una transmisión eficiente de la información a lo largo de la red, pues la disposición de los nodos en la estructura central favorece la distribución de la información a todos los nodos de la red.
3. *Disasortativa.* Las redes *disasortativas* se caracterizan por el hecho de que las conexiones se concentran entre nodos de distintos módulos. Por tanto, esta distribución favorece la transmisión de información entre las distintas comunidades de la red.

Podemos observar una diferencia funcional entre las estructuras *asortativas* y no *asortativas*, pues mientras las primeras se basan en la especialización de las comunidades, las

segundas procuran una mejor distribución de la información.

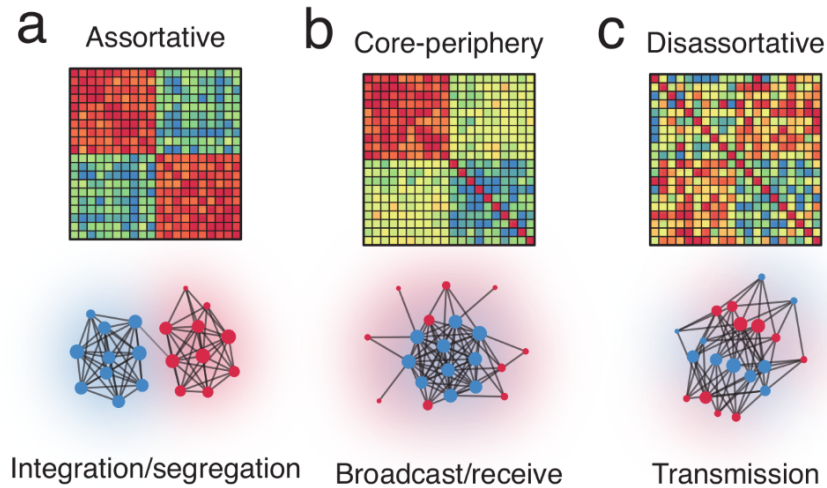


Figura 2.2: De izquierda a derecha: red asociada a una estructura asortativa, red del tipo centro-periferia y red con estructura disasortativa.

La Figura 2.2, de [4], presenta tres redes distintas. La primera tiene una estructura comunitaria de tipo asortativa, con dos módulos con una alta densidad de conexiones interna y tan solo una conexión entre ambos módulos. Esta estructura da lugar a una matriz de adyacencia con dos bloques diagonales (representados de color rojo) que se asocian a los dos núcleos densamente conectados, mientras que el resto de la matriz nos indica que no hay demasiada relación entre ambos módulos. En la segunda tenemos de nuevo una red de dos módulos, pero en esta ocasión la estructura se corresponde a una del tipo centro-periferia. Podemos observar que de estas dos comunidades hay una conformada por nodos densamente conectados entre sí, mientras que la otra viene dada por nodos desconexos que se asocian con los vértices del núcleo. Este hecho implica que la estructura comunitaria de esta red determina, respectivamente, un núcleo y una periferia. Podemos asociar esta distribución a una matriz de adyacencia que se va difuminando progresivamente, simbolizando una gran cantidad de conexiones entre nodos del núcleo, unas cuantas aristas entre los dos módulos y una periferia desconexa. El último grafo posee una estructura disasortativa con dos módulos densamente conectados entre sí pero con una mínima densidad de conexiones dentro de las comunidades. Esta estructura da lugar a una matriz de adyacencia que expone una gran cantidad de conexiones entre los módulos y una relación mínima o nula entre los nodos de la misma comunidad.

## 2.5. Núcleo de una red

Finalmente, tratamos de determinar un conjunto de nodos que permite que la red, tanto en el conectoma de *C. elegans* como en las redes de referencia, cumpla su función correctamente. La idea intuitiva es que si eliminamos ese núcleo, entonces sería imposible transmitir información de unos nodos a otros y la red dejaría de realizar sus funciones. A este conjunto lo llamamos *núcleo* de la red.

A partir de esta idea, varios autores han estudiado distintos subconjuntos que pudieran jugar el papel de núcleo. Uno de ellos es el centro de la red, es decir, el conjunto dado por los nodos centrales de la red. Esto se debe a que estos nodos son los encargados de transmitir la información a través de la red de la forma más rápida y eficiente posible y si fuesen eliminados la información circularía lentamente e incluso, en algunos casos, se podría desconectar la red.

Un conjunto propuesto por numerosos autores como núcleo de las redes complejas es el *rich club*, formado por el conjunto de nodos con mayor grado de la red. El estudio del *rich club* del conectoma de *C. elegans* puede verse en [19]. Inicialmente este conjunto fue considerado como núcleo de la red debido a su importancia a nivel práctico pues al ser estos nodos los que acumulan la mayor parte de conexiones, se consideraba que estos actuaban como puente a lo largo de la red. Por tanto, si se prescindiera de este subconjunto, cabría esperar que la red perdiera eficacia a la hora de transmitir información y que su función pudiera verse afectada.

Más recientemente, se ha propuesto como núcleo del conectoma de *C. elegans* un subconjunto de la red de 20 interneuronas, al que llaman *hourglass weist* o *cuello de botella* [16]. Estas interneuronas forman un cuello de botella atravesado por la mayoría de los caminos geodésicos que unen las neuronas sensoriales con las motoras. De esta forma, se argumenta que estas neuronas son las encargadas de transmitir los impulsos nerviosos, transformando los estímulos externos captados por las neuronas sensoriales en respuestas de las neuronas motoras.

Sin embargo, a menudo estos conjuntos no representan el núcleo funcional de la red en el sentido que hemos propuesto, pues aunque su eliminación suponga el fallo de ciertas funciones, podría mantenerse una conectividad, si no global, al menos de una gran componente, permitiendo el mantenimiento de una buena parte de las funciones de la red.

Por ello, en este trabajo tratamos de definir un núcleo funcional de la red a partir del estudio comparado del conectoma de *C. elegans* y las redes de control descritas antes. La idea de partida es considerar un subconjunto de nodos de modo que, si fuese eliminado, la red dejaría de funcionar correctamente, tanto globalmente como a pequeña o media escala. Para ello analizamos el efecto que produce sobre la topología del conectoma de *C. elegans*,

la red telemática Internet2 y de la red jerárquica de Ravasz la eliminación de los sucesivos conjuntos de vértices:

1. El centro de la red, encargado de transmitir la información de la forma más rápida posible.
2. La red de conectores, formada por los nodos con la mayor densidad de conexiones extracomunitarias.
3. La subred de conectores centrales, que combina ambas características.

Las propiedades de estas subredes están claramente vinculadas con la asortatividad y el tipo de estructura de la red [4] y la existencia de un cuello de botella [16], pero nuestras conclusiones muestran que las redes cerebrales poseen una arquitectura más compleja que la que proponen esas y otras referencias.



## Capítulo 3

# Resultados para el conectoma de *C. elegans*

En esta capítulo aplicamos los métodos para la detección de la estructura comunitaria al conectoma de *C. elegans* y tratamos de detectar su núcleo funcional.

### 3.1. Estructura comunitaria modular

Comenzamos estudiando la estructura comunitaria obtenida mediante el algoritmo de Lovaina en el conectoma de *C. elegans*. La Q-modularidad obtenida es de 0.414, lo que significa que la densidad de conexiones en los módulos dados por el algoritmo de Lovaina no es especialmente elevada, sino que la proporción de conexiones entre neuronas de distintos módulos es bastante alta en relación a las conexiones interiores de cada comunidad. Esta baja modularidad no implica un mal cálculo de la estructura comunitaria, sino que la red es tan compleja que no es posible dividirla en módulos bien definidos con una alta densidad de conexiones internas.

Por otra parte, la estructura comunitaria proporcionada por el algoritmo está constituida por cinco comunidades, donde el primer módulo está formado por 20 neuronas, el segundo por 58, el tercero por 92, el cuarto por 52 y el último por 78. La composición exacta de cada módulo se puede ver en la Tabla 3.1.

Módulo 1	Módulo 2	Módulo 3	Módulo 4	Módulo 5
I1L	ADFL	IL1DL	FLPL	AVM
I1R	ADFR	IL1DR	FLPR	PVM
I2L	ADLL	IL1L	RID	BDUR
I2R	ADLR	IL1R	AVAR	PVDL
I3	ASEL	IL1VL	AVBR	PVDR
I4	ASER	IL1VR	DA01	AQR
I5	ASGL	IL2DL	DA02	PQR
I6	ASGR	IL2DR	DA03	PDEL
M1	ASHL	IL2L	DA04	PDER
M2L	ASHR	IL2R	DA05	PHAL
M2R	ASIL	IL2VL	DA06	PHAR
M3L	ASIR	IL2VR	DB01	PHBL
M3R	ASJL	OLQDL	DB02	PHBR
M4	ASJR	OLQDR	DB03	PHCL
MI	ASKL	OLQVL	DB04	PHCR
NSML	ASKR	OLLL	DB05	AVFL
NSMR	AWAL	OLLR	AS01	AVFR
M5	AWAR	CEPDL	AS02	AVG
MCL	AWBL	CEPDR	AS03	AVHL
MCR	AWBR	CEPVL	AS04	AVHR
	AWCL	ALML	AS05	AVJL
	AWCR	ALNL	AS06	AVJR
	AFDL	ALNR	AS07	AVL
	AFDR	PLNL	DD01	DVA
	OLQVR	PLNR	DD02	DVB
	CEPVR	BDUL	DD03	DVC
	ALMR	SDQL	VA01	PVNL
	BAGL	SDQR	VA02	PVNR
	URXL	BAGR	VA03	PVPL
	URXR	URYDL	VA04	PVPR
	AUAL	URYDR	VA05	PVQR
	AUAR	URYVL	VA06	PVR
	ADER	URYVR	VA07	PVT
	AIAL	ADEL	VA08	PVWL
	AIAR	AIBR	VB01	PVWR
	AIBL	RIAL	VB02	PVCL
	AIYL	RIBL	VB03	PVCR
	AIYR	RIBR	VB04	AVAL
	AIZL	RICL	VB05	AVBL
	AIZR	RICR	VB06	AVDL
	ADAL	RIGL	VD01	AVDR
	ADAR	RIH	VD02	LUAL
	AIML	RIS	VD03	LUAR
	AIMR	RMFL	VD04	SABD
	AINL	RMFR	VD05	SABVL
	AINR	RMGL	VD06	SABVR
	ALA	URBL	VD07	DA07
	RIAR	RIPL	VC01	DA08

Módulo 1	Módulo 2	Módulo 3	Módulo 4	Módulo 5
	RIFL	RIPR	VC02	DA09
	RIFR	AVKL	VC03	PDA
	RIGR	AVKR	VC04	DB06
	RIR	AVEL	DD04	DB07
	RMGR	AVER		AS08
	URBR	RIML		AS09
	PVQL	RIMR		AS10
	SMBVR	RMDDL		AS11
	HSNL	RMDDR		PDB
	HSNR	RMDL		VA09
		RMDR		VA10
		RMDVL		VA11
		RMDVR		VA12
		RMED		VB07
		RMEL		VB08
		RMER		VB09
		RMEV		VB10
		RMHL		VB11
		RMHR		VD08
		RIVL		VD09
		RIVR		VD10
		URADL		VD11
		URADR		VD12
		URAVL		VD13
		URAVR		VC05
		SAADL		VC06
		SAADR		PLML
		SAAVL		PLMR
		SAAVR		DD05
		SIADL		DD06
		SIADR		
		SIAVL		
		SIAVR		
		SIBDL		
		SIBDR		
		SIBVL		
		SIBVR		
		SMBDL		
		SMBDR		
		SMBVL		
		SMDDL		
		SMDDR		
		SMDVL		
		SMDVR		

Tabla 3.1: Composición de cada módulo proporcionado por el algoritmo de Lovaina y distribución por ganglios.

Una vez aplicado el algoritmo conviene comparar la estructura comunitaria dada por este con la división anatómica del conectoma, comprobando si existe alguna relación entre ambas particiones. Al realizar la comparativa obtenemos que las 20 neuronas del primer módulo corresponden al ganglio de la faringe de *C. elegans*, por lo que en este caso la comunidad detectada por el algoritmo se corresponde estrictamente con uno de los ganglios del nematodo. Para observar la relación entre los restantes módulos y los ganglios de *C. elegans*, en la Figura 3.1 describimos cómo se distribuyen las neuronas por comunidades e indicamos mediante colores su distribución ganglionar. En esta gráfica no representamos el primer módulo, puesto que se corresponde perfectamente con el ganglio de la faringe.

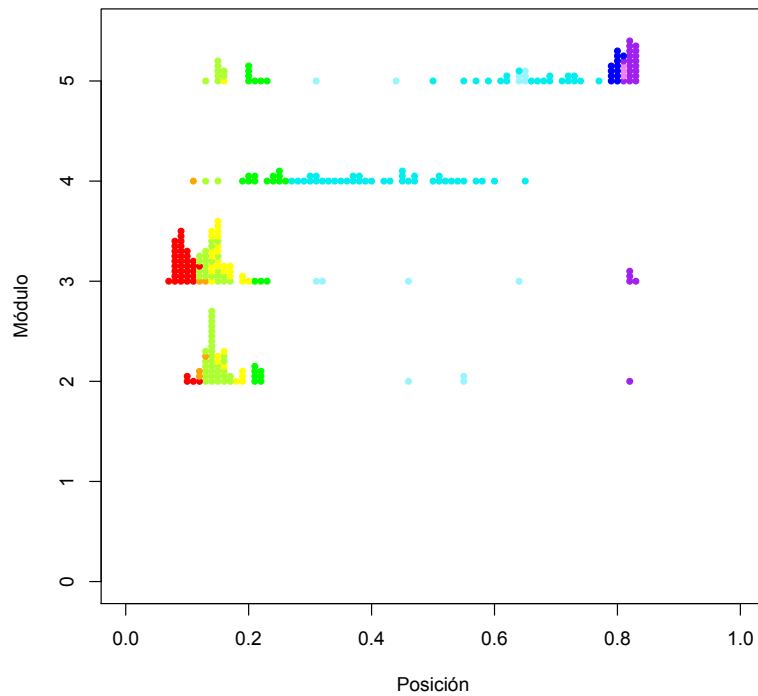


Figura 3.1: Representación de la composición de los diferentes módulos en función de la distancia de cada neurona a la cabeza del nematodo. Los colores indican la pertenencia de cada neurona a un ganglio según la Tabla 3.2.

Podemos observar que la mayoría de las neuronas que conforman las comunidades 2 y 3 están ubicadas en la cabeza de *C. elegans*, mientras que las del cuarto módulo se distribuyen a lo largo del tronco del nematodo y las de la quinta comunidad se concentran en su cola. Además, como ya dijimos, las neuronas del primer módulo se corresponden íntegramente

Ganglio anterior	Ganglio dorsal	Ganglio ventral	Ganglio lateral	Ganglio retrovesicular
Rojo	Naranja	Amarillo	Lima	Verde
Grupo ventral central	Ganglio posterolateral	Ganglio preanal	Ganglio dorsorectal	Ganglio lumbar
Azul claro	Celeste	Azul	Violeta	Morado

Tabla 3.2: Relación entre colores y ganglios en la Figura 3.1.

con la faringe, ubicada en la cabeza de *C. elegans*.

Con todo, a pesar de esta relación espacial de las neuronas que conforman los módulos, no parece que haya una relación determinante entre ganglios y módulos. En cualquiera de las comunidades representadas en la Figura 3.1 nos encontramos neuronas de ganglios distintos. Por ejemplo, en la segunda y en la tercera comunidad tenemos neuronas asociadas al ganglio anterior, ubicadas en la cabeza del nematodo y también neuronas del ganglio lumbar, que están en la cola. Esto significa que, pese a la particular distribución espacial del conectoma, no hay una relación directa entre ganglios y conjuntos de neuronas altamente conectadas, por lo que el hecho de que dos neuronas estén posicionalmente próximas no implica que estén conectadas entre sí.

Por otra parte, los casos de las neuronas de los ganglios lateral y retrovesicular tienen especial interés, ya que están repartidas en todos los módulos representados en la Figura 3.1. Este hecho es una prueba más de que no hay una razón posicional en la estructura comunitaria dada por el algoritmo de Lovaina. De hecho, estudiando el conectoma podemos observar que hay ganglios disconexos, por lo que difícilmente el algoritmo los agruparía en el mismo módulo.

De esta forma, podemos afirmar que en vista a los resultados obtenidos, no existe una relación directa entre la estructura modular dada por el algoritmo de Lovaina y la agrupación anatómica en ganglios presente en *C. elegans*. Sin embargo, podemos plantearnos si existe una relación entre agrupación modular y espacial, vinculada a algún tipo de función específica, a una escala más reducida. Por ello, aplicamos el algoritmo de Lovaina a cada una de las comunidades, es decir, repetimos el proceso de cálculo de la estructura comunitaria para cada uno de los 5 módulos detectados inicialmente.

Así, mediante la nueva aplicación del algoritmo, obtenemos el reparto en subcomunidades representado en la Figura 3.2.

Según los nuevos cálculos, tenemos que el primer y el segundo módulo se separan en tres submódulos, mientras que el tercero, el cuarto y el quinto se separan en cuatro. Cada módulo se representa por las neuronas comprendidas entre dos líneas continuas y las líneas punteadas separan los submódulos obtenidos en cada comunidad. En esta representación

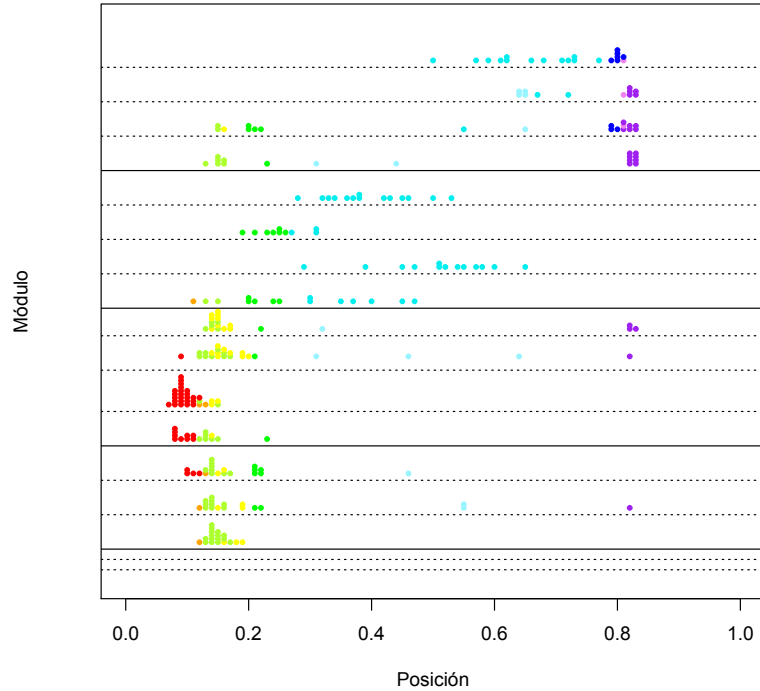


Figura 3.2: Módulos desagregados en comunidades más pequeñas manteniendo la posición y el color de pertenencia a cada ganglio.

volvemos a obviar las neuronas del primer módulo, pues como dijimos se corresponden con el ganglio de la faringe. Para el resto de las comunidades observamos que persisten los problemas que aparecían a escala global, pues algunos submódulos combinan neuronas de ganglios completamente distintos, aunque como cabía esperar la uniformidad espacial es mayor. Además, aunque hay submódulos con neuronas de un único ganglio, este se encuentra repartido entre diferentes submódulos.

En resumen, hemos comprobado que no hay una relación aparente entre la posición y la estructura comunitaria de las neuronas de *C. elegans*, ni a escala global, ni a escala más reducida. De todas formas, podríamos preguntarnos si la conexión entre las neuronas obedece a razones funcionales más que espaciales. Ahora estudiamos si las comunidades obtenidas con el algoritmo de Lovaina se corresponden con la división funcional del conectoma. Para ello, en la Figura 3.3, mostramos la composición de cada módulo indicando la posición y la función de cada neurona. Obviamos de nuevo las neuronas asociadas a la faringe, pues desempeñan una función independiente por sí mismas.

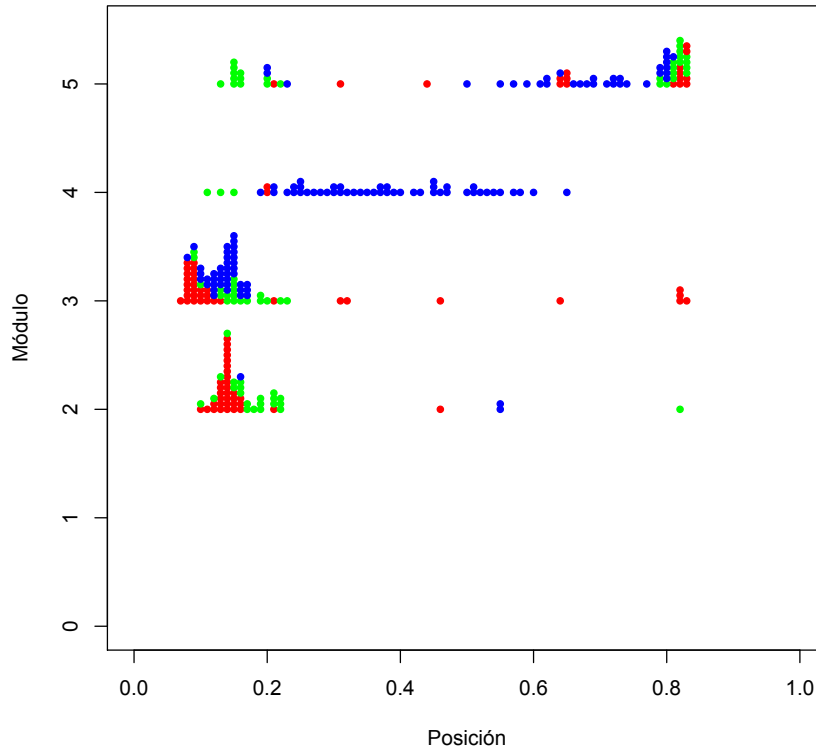


Figura 3.3: Representación de la composición de los diferentes módulos según de la función de cada neurona marcada con un color diferente de acuerdo con la Tabla 3.3.

En la siguiente tabla puede verse el código de colores asociado a los diferentes tipos de neuronas:

Neuronas sensoriales	Interneuronas	Neuronas motoras
Rojo	Verde	Azul

Tabla 3.3: Relación entre colores de la Figura 3.3 y funciones.

En la gráfica podemos observar de nuevo que, a priori, no hay relación entre los módulos detectados por el algoritmo de Lovaina y las funciones asignadas a las neuronas. En todos los módulos, a excepción del relacionado con la faringe, podemos encontrar neuronas que desempeñan funciones motoras, sensoriales o que actúan como interneuronas, lo que indica que a escala global no se puede afirmar que exista una relación directa entre los módulos y

las funciones realizadas por las neuronas que los conforman. Esto implica que, aunque dos neuronas desempeñen una misma función, no tienen más probabilidades de estar conectadas entre sí.

Con todo, igual que al comparar la estructura comunitaria y la partición anatómica, cabría preguntarse si la relación entre funciones y módulos, que no se da a nivel global, puede surgir a menor nivel. Para ello, igual que hicimos antes, volvemos a aplicar el algoritmo de Lovaina a cada uno de los módulos obtenidos inicialmente y vemos cómo están distribuidas las neuronas de cada submódulo según su función.

Volvemos a representar los submódulos en la Figura 3.4 y observamos que se mantiene el problema, ya que a pesar de que en la cuarta comunidad hay submódulos formados solo por neuronas motoras, este tipo de neurocitos aparecen en otros submódulos. Esto significa que, obviando la faringe, no parece haber relación entre los módulos y bloques funcionales en el conectoma de *C. elegans*. Por tanto, podemos afirmar que las neuronas encargadas de ejecutar funciones comunes no forman una subred densamente conectada.

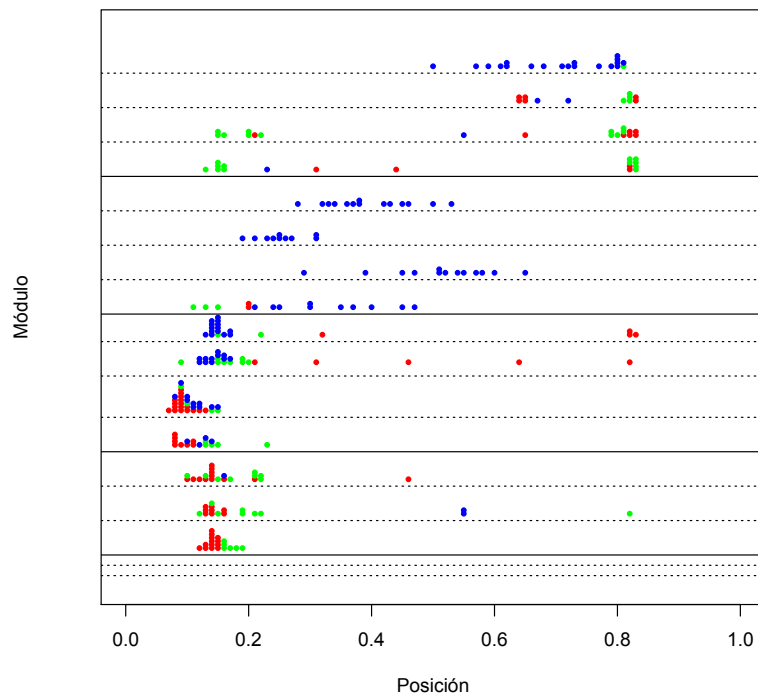


Figura 3.4: Módulos desagregados en comunidades más pequeñas manteniendo la posición y la función de cada neurona.

## 3.2. Modelos de bloques estocásticos

Hemos comprobado que la estructura comunitaria que maximiza la modularidad de la red no guarda relación ni con la distribución de las neuronas en ganglios ni con la división funcional del conectoma. Cabría preguntarse ahora si los ganglios o la distribución funcional tienen relación con los SBM, de modo que las neuronas que conforman los módulos siguen patrones de conectividad similares entre sí.

En este apartado implementamos los métodos de bloques estocásticos para la detección de estructura comunitaria del conectoma de *C. elegans*. En principio, en base a los resultados obtenidos tras aplicar el algoritmo de Lovaina, pedimos que los modelos de bloques estocásticos nos devuelvan una estructura comunitaria formada por 5 módulos.

Obtenemos 5 comunidades de 2, 20, 136, 96 y 46 nodos, donde el segundo módulo se corresponde con el ganglio de la faringe, por lo que al igual que ocurría con el algoritmo de Lovaina, los SBM son capaces de detectar este ganglio. En la Figura 3.5 representamos la partición obtenida mediante SBM. Como antes, mostramos la relación entre los módulos estocásticos y la posición de las neuronas que los conforman y al mismo tiempo comparamos la estructura comunitaria con la división anatómica de *C. elegans*, indicando la pertenencia a los ganglios con los colores que aparecen en la Tabla 3.2. En esta figura, al igual que ocurría con el algoritmo de Lovaina, no representamos el segundo módulo, asociado íntegramente a la faringe, pues esta relación ya justifica su interpretación y no es necesario estudiarlo junto al resto del conectoma.

La relación de colores y ganglios de la Figura 3.5 se mantiene de la Tabla 3.2.

Podemos observar un reparto muy asimétrico, con un módulo que tan solo cuenta con dos neuronas frente a otro que tiene 136. Esto denota unos patrones de conectividad muy variados entre las neuronas del conectoma, pues el hecho de que un módulo lo conformen tan solo dos neurocitos indica que la forma en que estas dos neuronas interactúan con el resto del conectoma no se parece a las interacciones presentes en el resto de la red. De hecho, estas dos neuronas son las encargadas de conectar el ganglio de la faringe con el resto del conectoma, por lo que sus patrones de conexión son distintos a los de otras neuronas de la red.

Al igual que ocurría previamente, volvemos a observar que no hay una relación directa entre ganglios y módulos puesto que, pese a agrupar la faringe y el ganglio dorsal en una única comunidad cada uno, el algoritmo divide el resto de los ganglios en varios módulos, por lo que no parece que el algoritmo detecte una relación entre conjuntos con similares patrones de conectividad y la estructura anatómica del nemátodo. Sin embargo, en la Figura 3.5, sí que podemos observar una ligera relación entre la posición y los módulos obtenidos

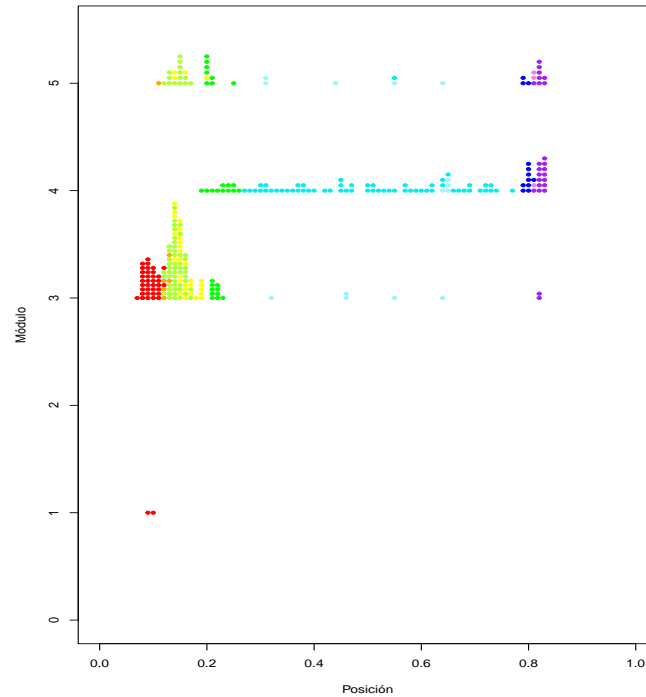


Figura 3.5: Representación de la composición de los diferentes módulos en función de la distancia de cada neurona a la cabeza del nematodo. Los colores indican la pertenencia de cada neurona a un ganglio según la Tabla 3.2.

con los SBM, pues a excepción de la quinta comunidad, el resto se puede corresponder con una posición anatómica particular de cada módulo. El primero, que tan solo cuenta con dos neuronas se relaciona con la cabeza de *C. elegans*, el segundo, como dijimos, con la faringe, el tercero con la cabeza y la parte superior de tronco y la cuarta con la cola y la parte media y baja del tronco del nematodo. La quinta no parece tener una posición específica en *C. elegans*, sino que tiene neuronas a lo largo de todo el cuerpo de *C. elegans*. Esta relación espacial nos permite afirmar que aunque SBM no detecta una relación entre ganglios y los módulos obtenidos, sí que detecta ligeramente cómo se distribuyen los neurocitos a lo largo del conectoma.

Por último, de nuevo, tratamos de buscar una relación entre la función que realizan las neuronas del conectoma y los módulos propuestos por los modelos de bloques estocásticos. Para ello mostramos una representación análoga a la de la Figura 3.3 con los resultados de los SBM.

A la vista de los resultados obtenidos en la Figura 3.6 podemos afirmar que hay una ligera relación entre comunidades y función, pues la mayoría de neuronas del tercer módulo

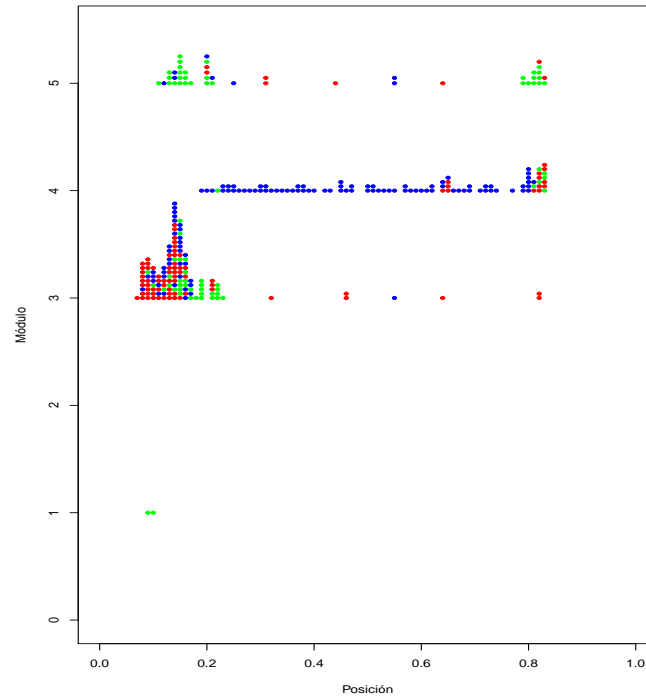


Figura 3.6: Representación de la composición de los diferentes módulos según de la función de cada neurona marcada con un color diferente de acuerdo con la Tabla 3.3.

son sensoriales, las del cuarto motoras y las del quinto interneuronas. Aún con todo, esta relación no es perfecta, pues en cada módulo podemos encontrar neuronas con todo tipo de funciones. Esto se debe a la complejidad del conectoma, donde las sinapsis de control difuminan en algunos momentos los patrones de conectividad.

Por tanto, aunque los SBM no detectan una relación entre ganglios y módulos más allá de la faringe, sí que detecta una ligera vinculación funcional de los módulos. Esto significa que neuronas del mismo tipo tienden a relacionarse con el resto del conectoma de una forma similar. Además la aparición del primer módulo, conformado por tan solo dos neuronas se corresponde con el hecho de que estas son las únicas encargadas de conectar el ganglio de la faringe al resto del conectoma, por lo que sus patrones de conectividad no se parecen a los del resto de neuronas. Hay que señalar que algunos autores han ampliado el número de módulos, tanto usando el algoritmo de Lovaina [17] como los SBM [13], para intentar ajustar la estructura comunitaria con funciones más específicas, aunque los resultados distan de ser concluyentes.

Como comentario final sobre el estudio del conectoma de *C. elegans*, conviene indicar que parece más justificada una división comunitaria como la descrita mediante los SBM

que la proporcionada por el algoritmo de Lovaina, pues aunque el primer método plantea dudas sobre los patrones de conexión internos, podemos encontrar una cierta vinculación de la partición con aspectos funcionales.

### 3.3. Centro y núcleo

Por último, realizamos un estudio individual de las neuronas según su coeficiente de participación y de centralidad. Con ello tratamos de detectar un núcleo funcional de la red que contenga aquellas neuronas que permiten el correcto funcionamiento del conectoma de *C. elegans*.

En primer lugar estudiamos el coeficiente de participación de los nodos según el procedimiento descrito en el Capítulo 2. Recordemos que las neuronas se consideran conectoras cuando su coeficiente de participación es mayor o igual que 0.3. En la Figura 3.7 representamos en orden creciente los valores obtenidos, pudiendo observar así claramente la proporción de neuronas conectoras y provinciales.

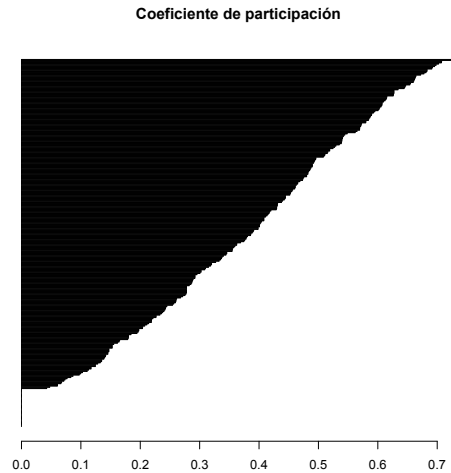


Figura 3.7: Representación del coeficiente de participación sobre las neuronas del conectoma. La disposición de sus valores es progresiva, de menos a más.

La Figura 3.7 muestra que algo más de la mitad de las neuronas son conectoras y las restantes son provinciales. Esto nos indica que los módulos de *C. elegans* son difusos, pues los elevados valores del coeficiente de participación indican que neuronas de un módulo se conectan frecuentemente con las de otras comunidades. Esto se relaciona con el valor relativamente bajo de la modularidad  $Q = 0,414$ . Particularizando, podemos señalar que hay

175 neuronas conectoras frente a las restantes 125 que son provinciales. Además, debido a su interés, estudiamos como varía la disposición de los valores del coeficiente de participación en cada módulo, observando cuales son los responsables de distribuir la información entre las comunidades. Presentamos los resultados de cada comunidad en la Figura 3.8.

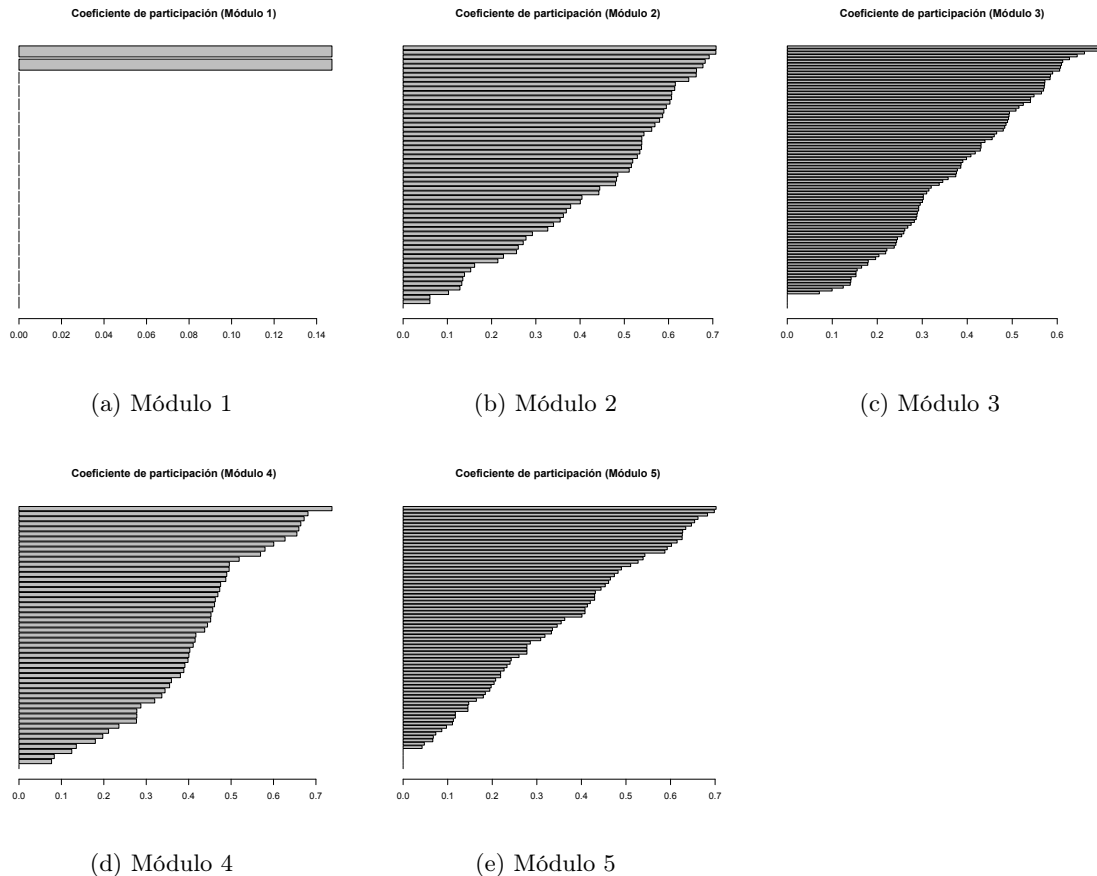


Figura 3.8: Distribución progresiva de los coeficientes de participación por módulo. De izquierda a derecha y de arriba a abajo: Módulo 1, módulo 2, módulo 3, módulo 4, módulo 5.

En la segunda, tercera, cuarta y quinta comunidad, el coeficiente de participación muestra el mismo fenómeno global: hay multitud de nodos conectoras en relación a las neuronas totales de cada comunidad. Sin embargo, en el primer módulo podemos observar que de 20 nodos, 18 tienen un coeficiente de participación nulo, y los de los otros dos toman un valor de alrededor de 0.14. Esto lo podemos entender del siguiente modo: las neuronas de la faringe están mayoritariamente agrupadas entre sí, habiendo solo dos que tienen unas pocas conexiones con neurocitos de otras comunidades. Esto nos indica que la faringe funciona de modo efectivo como un módulo independiente, lo que se asocia a la

realización de una función independiente del resto del conectoma. La faringe representa un conjunto eficiente e independiente que transmite la información, principalmente, a nivel local. Al contrario, para los otros cuatro módulos, observamos conjuntos difusos con bastantes conexiones intercomunitarias. Estos elevados valores del coeficiente de participación pueden justificar la incapacidad del algoritmo de Lovaina para detectar las particiones anatómicas o funcionales más allá de la faringe, pues estos elevados valores denotan que la red es tremendamente compleja y el hecho de proponer una estructura comunitaria que maximice la modularidad no implica que sus módulos posean ninguna relación estructural o funcional.

En la Figura 3.9 representamos los coeficientes de centralidad, de la misma manera que hicimos con los de participación.

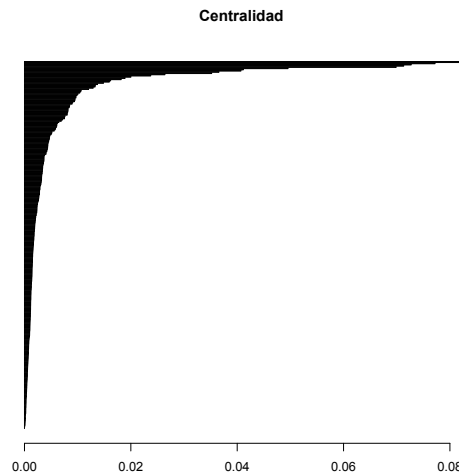


Figura 3.9: Representación del coeficiente de centralidad sobre las neuronas del conectoma. La disposición de sus valores es progresiva, de menos a más.

En esta gráfica podemos observar que el coeficiente de centralidad aumenta exponencialmente, lo que sugiere que la información que fluye entre dos neuronas apenas circula por el resto del conectoma, sino que la red propicia que las conexiones entre neuronas sean prácticamente directas. De hecho, la distancia media entre dos neuronas del conectoma es inferior a 4 sinapsis, lo que indica que prácticamente no es necesaria la utilización de neuronas puente para enviar información a lo largo del conectoma. Observando la gráfica podemos afirmar que no hay ningún nodo cuyo coeficiente de centralidad supere el nivel de 0.1. Esto implica que, según el criterio de Sporns et al. [18], en el conectoma de *C. elegans* no hay nodos centrales, es decir, no hay nodos que funcionen como paso necesario de las conexiones a través de la red. Esto lo podemos asociar con la especialización de las neuronas,

donde cada célula realiza una función particular especializada y luego estas se agrupan entre sí en el resto del conectoma, necesitando una amplia cantidad de conexiones a media escala. Esta complejidad da lugar a valores de centralidad más bajos de lo habitual y por ello, en base a este criterio, no es posible definir el centro de la red.

Sin embargo, cabe destacar que disminuyendo el nivel crítico del coeficiente de centralidad a la media de los valores obtenidos, sí que podríamos definir un centro formado por 60 neuronas. Estas, que son las de mayor centralidad, serían las principales encargadas de transmitir la información a lo largo de todo el conectoma. Con todo, no hay que olvidar que estos valores son relativamente bajos, pues el límite se toma en 0.005, por lo que no podemos considerar este conjunto como un centro efectivo, pues en multitud de ocasiones no es necesario tomar sus neuronas como puente para la transmisión de información.

A continuación representamos la distribución del coeficiente de centralidad según los módulos obtenidos con el algoritmo de Lovaina. Esto lo hacemos con el fin de detectar qué neuronas actúan como puente en cada comunidad y de observar qué módulos tienen más importancia a la hora de distribuir la información a lo largo del conectoma.

En la Figura 3.10 podemos observar que, salvo en el segundo módulo, se mantiene el fenómeno del crecimiento exponencial del coeficiente de participación a lo largo de la red. Por tanto, en la primera, tercera, cuarta y quinta comunidad podemos encontrar una o dos neuronas con una centralidad especialmente elevada, lo que se correspondería con las neuronas encargadas de transmitir la información de una forma eficiente en sus respectivos módulos. Por otra parte, en la segunda comunidad, podemos ver que no hay ningún nodo con un coeficiente de centralidad con un valor superior a 0.015 y que además, el crecimiento de los valores del coeficiente de centralidad no es tan pronunciado como en el resto de módulos. Esto lo podemos achacar a que, en mayor o menor medida, las neuronas del segundo módulo no cuentan con una gran cantidad de sinapsis que transmitan impulsos a lo largo del conectoma, sino que estas buscan controlar el correcto funcionamiento de la red y que no se de una disfuncionalidad del conectoma si alguna de las neuronas centrales deja de trabajar correctamente.

En base a los resultados obtenidos cabría preguntarse si el centro obtenido se corresponde con un núcleo funcional de la red, es decir, si al eliminar este conjunto de neuronas del conectoma la red dejaría de funcionar tanto globalmente como a nivel local. En caso de suprimirlo obtendríamos que la red resultante se separaría en dos componentes: una tendría 17 nodos y la otra 223. Este resultado nos indica que la red no se disocia en pequeñas componentes que imposibilitan la coordinación neuronal de *C. elegans*, sino que se mantiene una gran componente que seguiría coordinando parte de las funciones de *C. elegans* a nivel semilocal. Este hecho es una muestra de la complejidad de la red y de su diseño robusto

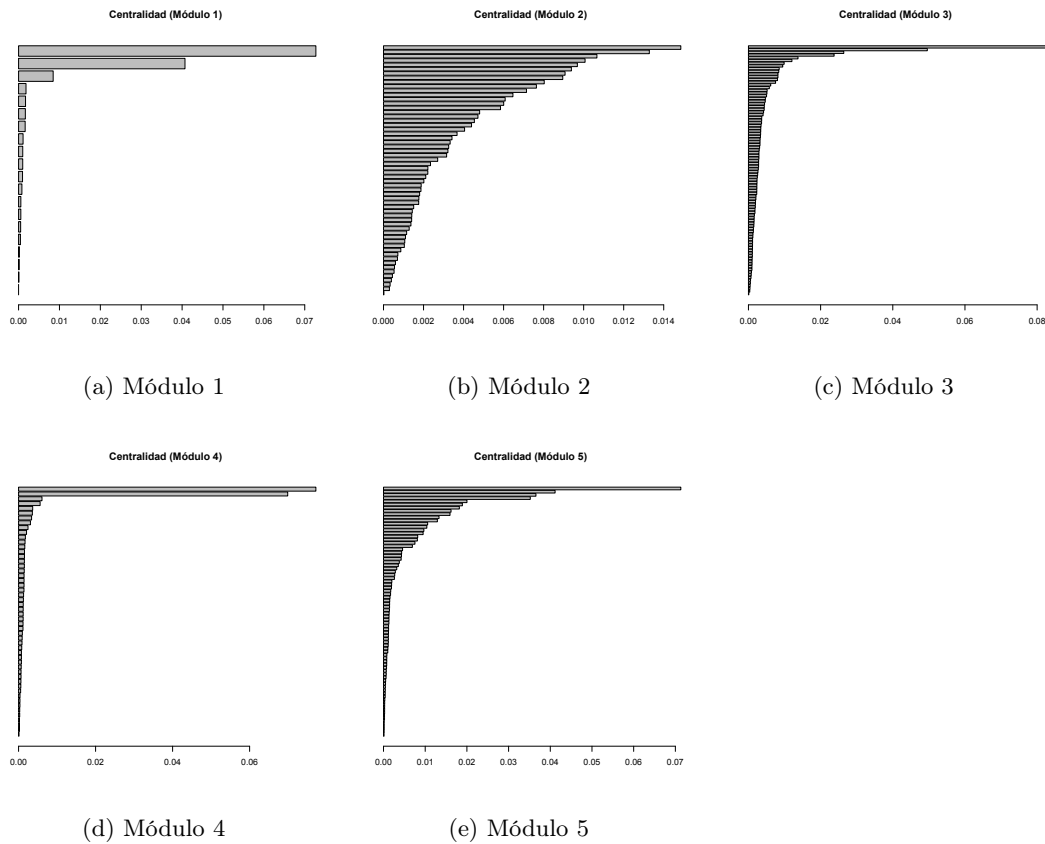


Figura 3.10: Distribución progresiva de los coeficientes de centralidad por módulo.

frente a perturbaciones, en este caso percolación de nodos, evitando así la desaparición del nematodo por una posible mutación. Además, cabe destacar que la componente conformada por 17 neuronas se relaciona con el ganglio de la faringe, que se separaría por completo del resto del conectoma realizando una función individual. Por tanto no parece razonable aceptar que el centro del conectoma actúe como núcleo funcional.

E. K. Towlson y sus coautores habían propuesto en [19] identificar al conjunto de neuronas de mayor grado, llamado *rich club*, como núcleo del conectoma. Este vendría dado por 11 neuronas somáticas que reunían gran parte de las conexiones presentes en el conectoma de forma que si estas fueran suprimidas la red perdería parte de su eficiencia. Posteriormente, K. M. Sabrin y sus coautores han propuesto en [16] considerar un núcleo más sofisticado que tuviera en cuenta la función de las neuronas en la totalidad del conectoma de modo que el hecho de prescindir de estas pudiera comprometer el funcionamiento de la red. Para ello, decidieron tomar como núcleo un conjunto de 20 interneuronas encargadas de actuar como puente en la transmisión de estímulos a lo largo del conectoma, transmitiendo

la información captada por las neuronas sensoriales a las motoras. Este subconjunto se llama *hourglass weist* o *cuello de botella*. En dicho artículo se argumenta que la supresión de ese conjunto de interneuronas supondría una transmisión ineficaz de la información a las neuronas motoras, dificultando la respuesta de *C. elegans* a estímulos externos. Sin embargo, tanto las 11 neuronas del *rich club*, como 18 de las 20 neuronas del *hourglass weist* se integran en el centro del conectoma y este es demasiado complejo para que se desconecte por completo al suprimirlos. Por ello tiene interés definir un subconjunto de mayor tamaño que actúe verdaderamente como un núcleo funcional de la red. Estudiemos los dos subconjuntos definidos en el Capítulo 2 como candidatos a núcleo.

Como proponíamos antes, nos interesamos a continuación por las neuronas que son al mismo tiempo conectoras, enviando información entre módulos, y centrales, distribuyéndola rápidamente. Como es un subconjunto del centro de la red, tampoco puede actuar como núcleo funcional del conectoma. Al hacer los cálculos correspondientes, vemos que este conjunto está compuesto por 53 neuronas somáticas. Al suprimir este subconjunto de la red obtenemos dos componentes conexas: una de 20, que se asocia al ganglio de la faringe, y otra de 227. Al igual que pasaba al considerar el centro, la red no se desconecta completamente, sino que queda una gran componente que seguiría ejecutando parte de las funciones del conectoma a nivel local.

Por último, si pensamos en la red de conectores como candidato a núcleo, vemos que dicha subred está formada por 254 neuronas, lo que genera un núcleo difuso y tremendamente amplio. En este caso suprimir este conjunto supone quedarse con 46 neuronas distribuidas en 6 componentes conexas de tamaño 20, 17, 4, 3 y dos neuronas aisladas. En este caso la componente de mayor tamaño se corresponde con las 20 neuronas de la faringe, que siguen actuando de forma independiente del resto del conectoma. Sin embargo, la supresión de la red de conectores hace inviable la transmisión de información en la red. Cabe destacar que esta subred contiene a todas las neuronas que conforman el *rich club* y también las del *hourglass weist*. Esto tiene una interpretación sencilla, ya que las neuronas de mayor valencia y las que actúan como puente al transmitir información entre neuronas sensoriales y motoras son una parte importante de la red, pero no imprescindible, luego por sí solas no conforman un núcleo funcional, pero sí forman parte de él.

Este estudio confirma lo que cabría esperar: debido a la complejidad de las redes neuronales de los seres vivos, en este caso de *C. elegans*, no parece posible definir un núcleo funcional de tamaño reducido de neuronas. Por razones evolutivas, una red biológica no puede depender de unas pocas neuronas, sino que es necesaria una arquitectura compleja que convierta al conectoma en una red robusta frente a la invasión de patógenos o las mutaciones (como se propone en [1, 2]).



## Capítulo 4

# Resultados para las redes de control

### 4.1. Internet2

#### 4.1.1. Algoritmo de Lovaina

Presentamos, en primer lugar, los resultados obtenidos al aplicar el algoritmo de Lovaina a la red telemática Internet2.

Como dijimos previamente, Internet2 es una red de fibra óptica gestionada por un consorcio de instituciones académicas estadounidenses que colaboran en investigación y educación. Esta red se caracteriza por una estructura jerárquica con 22 conectores primarios y 38 conectores secundarios, que forman un núcleo central por el que pasa toda la información y al que se conectan las restantes 251 instituciones académicas. La disposición de este núcleo central se puede observar en la Figura 1.3. Esta red fue estudiada en [1] como parte de un estudio de la robustez frente a los procesos de invasión de redes biológicas y tecnológicas, aunque su germen, conocido como Abilene, había aparecido en estudios previos en un contexto diferente [6]. Un estudio detallado del proceso de invasión fue realizado por J. Piñeiro en [14]. En ambos casos, los datos usados corresponden a abril de 2013.

Cabe destacar que los nodos que conforman el núcleo geográfico se asocian entre sí dando lugar a varios ciclos. Este hecho hace que sea lógico pensar que la estructura comunitaria de la red agrupe los nodos por ciclos, decidiendo mediante algún criterio en qué módulo añadir los centros universitarios que pertenecen a dos ciclos contiguos. Por tanto, sería esperable obtener una estructura comunitaria formada por seis módulos que agruparan cada ciclo en una comunidad.

Sin embargo, aplicando el algoritmo de Lovaina, obtenemos la estructura comunitaria que se describe en la Figura 4.1 y se detalla posteriormente en la Tabla 4.1.

En la Figura 4.1 podemos ver que, finalmente, no se da el reparto en comunidades

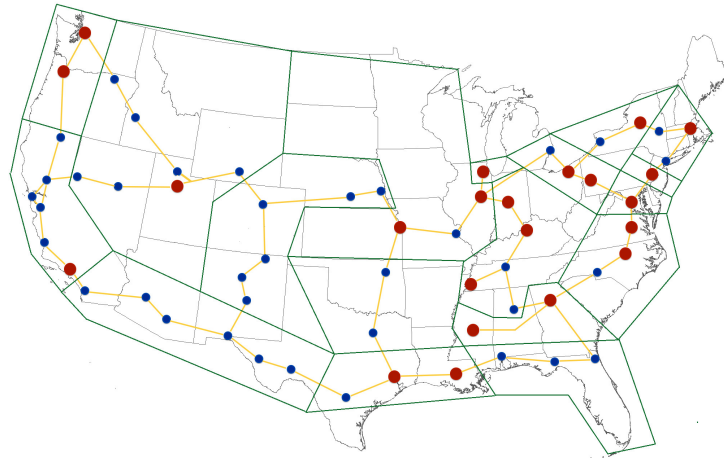


Figura 4.1: Representación de los módulos dados por el algoritmo de Lovaina al aplicarlo sobre la red asociada al consorcio de Internet2.

esperado, sino que los ciclos aparecen a menudo separados en distintos módulos. En este caso, el algoritmo de Lovaina nos devuelve 18 comunidades distintas y una  $Q$ -modularidad de 0.866. El algoritmo muestra una distribución particular de las comunidades a lo largo de los ciclos, justificada por un lado por la gran cantidad de nodos terminales de la red, que aumentan la densidad de conexiones de los módulos sin aportar conexiones externas con otras comunidades, y por otro a un efecto característico del algoritmo que se comenta más adelante. En la Tabla 4.1 detallamos la partición comunitaria completa dada por el algoritmo de Lovaina. En ella indicamos las universidades presentes en cada módulo y añadimos al lado de cada conector primario o secundario su valencia, es decir, el número de conexiones que tienen con otras universidades, eligiendo como conector principal de un módulo al de mayor valencia. Nótese que las instituciones que no están acompañadas de su valencia son nodos terminales de la red, por lo que su valencia es 1.

Conector principal	Módulo
MAX - 25	MAX = Mid-Atlantic Crossroads - 25 American University Catholic University of America Federal Highway Administration, Turner-Fairbank Highway Research Center George Washington University (The) Georgetown University Howard Hughes Medical Institute Johns Hopkins University Laboratory for Telecommunications Sciences (LTS) Library of Congress Morgan State University NASA Goddard Space Flight Center National Archives and Records Administration National Institutes of Health National Library of Medicine National Science Foundation NIST (National Institute of Standards and Technology) NOAA (National Oceanic & Atmospheric Administration, Washington, D.C.) Northrop Grumman Corporation United States Department of Agriculture, Agricultural Research Services (USDA) University of Maryland Baltimore University of Maryland Baltimore County University of Maryland College Park

---

OARnet - 13	<p>OARnet = Ohio Technology Consortium - 13  3ROX/Drexel = Three Rivers Optical Exchange - 6  Detroit OADM - 6  Carnegie Mellon University  Case Western Reserve University  Cleveland Institute of Music  Kent State University Main Campus  Marshall University  Miami University-Oxford  Michigan State University  Ohio University -Main Campus  Penn State (The Pennsylvania State University)  The Ohio State University  University of Akron, Main Campus, The  University of Cincinnati Main Campus  University of Michigan-Ann Arbor  University of Pittsburgh, Pittsburgh Campus  University of Toledo  Wayne State University  West Virginia University  Wright State University</p>
FLR/SoX - 14	<p>FLR/SoX = The Southern Crossroads - 14  MISSION = Mississippi Optical Network - 5  Auburn University  Centers for Disease Control and Prevention  Emory University  Georgia Institute of Technology  Georgia State University  Jackson State University  Medical University of South Carolina  Mississippi State University  Oak Ridge National Laboratory  University of Georgia  University of Mississippi  University of South Carolina - Columbia  University of Southern Mississippi  University of Tennessee</p>
MCNC/C-Light - 9	<p>MCNC/C-Light = North Carolina Research and Education Network - 9  MARIA = Mid-Atlantic Research Infrastructure Alliance - 7  Charlotte OADM - 3  Clemson University  Duke University  East Carolina University  George Mason University  North Carolina State University  Old Dominion University  Red Hat, Inc.  University of North Carolina at Chapel Hill  University of North Carolina At Greensboro  University of Virginia  Virginia Commonwealth University  Virginia Polytechnic Institute and State University  Wake Forest University</p>
Jacksonville OADM - 10	<p>Jacksonville OADM - 10  Mobile OADM - 2  Tallahassee OADM - 4  Florida Agricultural and Mechanical University  Florida Atlantic University  Florida International University  Florida State University  New World Symphony  University of Central Florida  University of Florida  University of Miami  University of Puerto Rico  University of South Florida</p>
MAGPI - 13	<p>MAGPI = Mid-Atlantic Gigapop in Philadelphia for Internet2 - 13  Lehigh University  New Jersey Institute of Technology  Princeton University  Rutgers, The State University of New Jersey  Temple University  The Children's Hospital of Philadelphia  The Philadelphia Orchestra Association  University of Delaware  University of Medicine and Dentistry of New Jersey  University of Pennsylvania  JSTOR</p>

---

---

Dallas OADM - 10	<p>Dallas OADM - 10  Tulsa OADM - 10  Arkansas State University  Baylor University  Oklahoma State University- Main Campus  Southern Methodist University  Texas Christian University  Texas Tech University  University of Arkansas at Little Rock  University of Arkansas for Medical Sciences  University of Arkansas Main Campus  University of North Texas  University of Oklahoma Norman Campus  University of Texas at Arlington  University of Texas at Dallas  University of Texas Southwestern Medical Center at Dallas  University of Tulsa  Wichita State University</p>
MREN - 9	<p>MREN = Metropolitan Research and Education Network - 9  Bradley University  DePaul University  Loyola University Chicago  Northern Illinois University  Northwestern University  Southern Illinois University Carbondale  University of Chicago  University of Illinois at Chicago</p>
PNG - 10	<p>PNG = Pacific Northwest Gigapop - 10  OG = Oregon Gigapop - 5  Microsoft Research  Montana State University - Bozeman  Oregon Health &amp; Science University  Oregon State University  Portland State University  Portland State University  University of Alaska Fairbanks  University of Hawaii At Manoa  University of Oregon  University of Washington  Washington State University</p>
NOX - 22	<p>NOX = The Northern Crossroads - 22  Albany OADM -5  New York City OADM - 11  Berklee College of Music  Boston College  Boston University  Brandeis University  Columbia University  Dartmouth College  EBSCO Industries, Inc  EBSCO Information Services  ESPN, Inc  Harvard University  IBM Corporation  Manhattan School of Music  Massachusetts Institute of Technology (MIT)  New York University  Northeastern University  Rensselaer Polytechnic Institute  Stony Brook University, State University of New York  The City University of New York  Tufts University  University At Albany, State University of New York  University of Connecticut  University of Maine  University of Massachusetts  University of New Hampshire  University of Rhode Island  University of Vermont  Worcester Polytechnic Institute  Yale University  Yeshiva University  JSTOR</p>
NYSERNet - 7	<p>NYSERNet Inc. - 7  Buffalo OADM - 3  Binghamton University  Cornell University  Rochester Institute of Technology  Syracuse University  University of Rochester  University At Buffalo, The State University of New York</p>

---

---

CIC OmniPoP - 16	<p>CIC OmniPoP = Committee on Institutional Cooperation OmniPoP - 16  GPN = Great Plains Network - 9  St. Louis OADM - 5  Iowa State University  Kansas State University  Marquette University  Mayo Medical School College of Medicine  Missouri University of Science and Technology  North Dakota State University Main Campus  South Dakota School of Mines and Technology  South Dakota State University  University of Illinois at Urbana-Champaign  University of Iowa - 2  University of Kansas Main Campus  University of Minnesota-Twin Cities  University of Missouri - Columbia  University of Missouri - Kansas City  University of Missouri - Saint Louis  University of North Dakota Main Campus  University of South Dakota  University of Wisconsin-Madison  University of Wisconsin-Milwaukee</p>
<hr/>	
CENIC - 16	<p>CENIC = Corporation for Education Network Initiatives in California - 16  Tionesta OADM - 2  Sacramento OADM - 6  San Francisco OADM - 6  Sunnyvale OADM - 8  San Luis Obispo OADM - 3  California Institute of Technology  California Polytechnic State University-San Luis Obispo  California State University, Fullerton  California State University, Office of the Chancellor  California State University-East Bay  Claremont Colleges  Complete Genomics Inc.  HP  Jet Propulsion Lab  Naval Postgraduate School  Pepperdine University  Stanford University  UCAR/NCAR (University Corporation for Atmospheric Research)  University of California, Office of the President  University of California-Berkeley  University of California-Davis  University of California-Irvine  University of California-Los Angeles  University of California-Riverside  University of California-San Francisco  University of California-Santa Barbara  University of California-Santa Cruz  University of Nevada-Las Vegas  University of Southern California</p>
<hr/>	
Denver OADM - 11	<p>Rancho de la Fe OADM - 2  Albuquerque OADM - 3  Raton OADM - 2  Denver OADM - 11  Heartwell OADM - 2  Omaha OADM - 3  Colorado School of Mines  Colorado State University  United States Antarctic Program (USAP)  United States Department of Commerce Boulder Labs  University of Colorado at Boulder  University of Colorado Denver  University of Denver  University of Nebraska - Lincoln  New Mexico State University -Main Campus  University of Wyoming</p>
<hr/>	
El Paso OADM - 5	<p>San Diego - 4  Phoenix OADM - 4  Tucson OADM - 3  El Paso OADM - 5  Valentine OADM - 2  Sanderson OADM - 2  Arizona State University  New Mexico State University -Main Campus  Northern Arizona University  San Diego State University  University of Arizona  University of California-San Diego  University of Texas at El Paso</p>

---

LEARN - 10	LEARN = Lonestar Education and Research Network - 10 LONI = Louisiana Optical Network Initiative - 5 San Antonio OADM - 3 Baylor College of Medicine Louisiana State University Rice University Stephen F. Austin State University Texas A&M University Tulane University University of Houston University of New Orleans University of Texas at Austin University of Texas at San Antonio University of Texas Medical Branch At Galveston
UEN - 6	UEN = Utah Education Network - 6 Reno OADM - 4 Eureka OADM - 2 Rawlins OADM - 3 Ogden OADM - 4 Boise OADM - 5 Rieth OADM - 2 Brigham Young University Desert Research Institute Idaho State University The University of Montana University of Idaho University of Nevada, Reno University of Utah Utah State University JSTOR
Indiana GigaPoP - 8	Indiana GigaPoP - 8 KyRON = Kentucky Regional Optical Network - 4 UM = University of Memphis - 2 Nashville OADM - 4 Birmigham OADM- 5 C-SPAN Indiana University Purdue University Main Campus Saint Louis University University of Alabama at Birmingham University of Alabama in Huntsville University of Alabama, The University of Kentucky University of Louisville University of Memphis University of Notre Dame Vanderbilt University Washington University in St. Louis

Tabla 4.1: Composición de cada módulo proporcionado por el algoritmo de Lovaina.

En base a estos resultados podemos afirmar que la estructura comunitaria descrita por el algoritmo de Lovaina reconoce la situación geográfica de las universidades y explica su interrelación, pues es capaz de separar en dos módulos la costa oeste, tres el sur del país, cinco el centro del estado y únicamente necesita más agrupaciones en la costa este. No obstante, el algoritmo no detecta los ciclos que conforman la estructura central. Este hecho provoca que la partición modular obtenida no sea la que mejor se ajusta a la estructura real, pues aunque esta división maximiza la modularidad de la red, separa nodos con posiciones relativas idénticas y conectividades similares (como centros de estrellas de diferentes tamaños). Con esto nos referimos a que dos nodos contiguos interiores a un ciclo son idénticos en cuanto a conectividad interna se refiere, lo que hace esperable que la estructura comunitaria no encuentre diferencias entre ellos y por tanto no los separe en

distintos módulos.

En [7] se observa que el algoritmo de Lovaina tiende a separar los ciclos de una red agrupando los nodos en pares, en lugar de una distribución más lógica en base a la simetría de los ciclos. En dicho artículo se presenta una red de  $k$  cliques (es decir, grafos completos), con  $c$  nodos cada uno, interconectados mediante una única arista hasta formar un ciclo (si colapsamos cada clique). Entonces cabría esperar que la estructura comunitaria obtenida mediante el algoritmo de Lovaina agrupara cada clique en un único módulo, pues como es lógico cada uno realiza una función independiente a menor escala que luego se transmite al resto de la red. Sin embargo, esto solo ocurre para redes con valores de  $k$  suficientemente pequeños, pues agrupar dos cliques adyacentes en una única comunidad supone una variación en la  $Q$ -modularidad de

$$\Delta Q = \frac{1}{k[\binom{c}{2} + 1]} - 2k^{-2},$$

que es positivo cuando  $k > c(c - 1) + 2$ . Por tanto, para un valor de  $k$  suficientemente grande el algoritmo agrupa dos a dos los cliques presentes en un ciclo, contradiciendo la distribución más razonable.

Regresando a Internet2 y a pesar de sus limitaciones, el algoritmo de Lovaina distingue módulos con una alta densidad de conexiones que conforman una estructura comunitaria compatible con la disposición geográfica de las universidades y otras instituciones de investigación. Por tanto, pese a su tendencia a disociar los nodos de los ciclos, el algoritmo de Lovaina encuentra una estructura comunitaria aceptable.

#### 4.1.2. Modelos de bloques estocásticos

Por otra parte, cabe preguntarse si el algoritmo de SBM es capaz de detectar la estructura comunitaria propuesta inicialmente. Para comprobarlo, aplicamos los modelos de bloques estocásticos usando como número de comunidades la cantidad de ciclos presentes en el bloque central de la red, es decir, le pedimos al algoritmo SBM que busque una estructura comunitaria dada por 6 módulos. A continuación, en la Tabla 4.2 mostramos los resultados obtenidos. De nuevo escribimos los nodos de cada comunidad y, al igual que con el algoritmo de Lovaina, indicamos al lado de cada conector principal o secundario su grado. Como puede verse, el algoritmo SBM tan solo devuelve 5 comunidades no vacías.

Conector principal	Módulo
FLR/SOX - 14	<p> LONI = Louisiana Optical Network Initiative - 5  OARnet = Ohio Technology Consortium - 13  3ROX/Drexel = Three Rivers Optical Exchange - 6  MCNC/C-Light = North Carolina Research and Education Network - 9  FLR/SOX = The Southern Crossroads - 14  NYSERNet Inc. - 7  San Diego OADM - 4  El Paso OADM - 5  Valentine OADM - 2  Sanderson OADM - 2  Rancho de la FE OADM - 2  Denver OADM - 11  Rawlins OADM - 3  Tallahassee OADM - 4  Brandeis University  Brigham Young University  C-SPAN  California Institute of Technology  California Polytechnic State University-San Luis Obispo  California State University, Office of the Chancellor  California State University-East Bay  Carnegie Mellon University  Case Western Reserve University  Catholic University of America  Centers for Disease Control and Prevention  Claremont Colleges  Clemson University  Cleveland Institute of Music  Duke University  EBSCO Industries, Inc  Federal Highway Administration, Turner-Fairbank Highway Research Center  Harvard University  HP  IBM Corporation  Idaho State University  Indiana University  Iowa State University  Jackson State University  JSTOR  Kansas State University  Kent State University Main Campus  Laboratory for Telecommunications Sciences (LTS)  Marshall University  Mayo Medical School College of Medicine  Medical University of South Carolina  NOAA (National Oceanic &amp; Atmospheric Administration, Washington, D.C.)  North Carolina State University  Northeastern University  Northern Arizona University  Ohio University -Main Campus  Stephen F. Austin State University  Stony Brook University, State University of New York  University of Medicine and Dentistry of New Jersey  University of New Hampshire  University of New Mexico Main Campus  University of New Orleans  University of North Carolina At Greensboro  University of North Dakota Main Campus  University of Pennsylvania  University of South Florida  University of Southern California  University of Southern Mississippi  University of Tennessee  University of Texas at Arlington  University of Texas at Dallas  University of Virginia  University of Wisconsin-Madison  University of Wisconsin-Milwaukee  University of Wyoming  Virginia Polytechnic Institute and State University  Wake Forest University  Washington State University  Washington University in St. Louis  Wayne State University  JSTOR </p>

---

Dallas OADM - 10	Heartwell OADM - 2 Dallas OADM - 10 St. Louis OADM - 5 Detroit OADM - 6 ESPN, Inc Johns Hopkins University University of Maryland Baltimore
------------------	---

---

CIC OmniPoP - 16	CIC OmniPoP = Committee on Institutional Cooperation OmniPoP - 16 UM = University of Memphis - 2 KyRON = Kentucky Regional Optical Network - 4 MAGPI = Mid-Atlantic Gigapop in Philadelphia for Internet2 - 13 Phoenix OADM - 4 Tucson OADM - 3 San Antonio OADM - 3 Albuquerque OADM - 3 Raton OADM - 2 Mobile OADM - 2 American University Arizona State University Boston University Bradley University California State University, Fullerton Colorado School of Mines East Carolina University Howard Hughes Medical Institute Marquette University Massachusetts Institute of Technology (MIT) NIST (National Institute of Standards and Technology) North Dakota State University Main Campus Penn State (The Pennsylvania State University) Texas A&M University University of North Carolina at Chapel Hill University of South Dakota University of Utah University of Washington
------------------	--

---

MAX - 25	OG = Oregon Gigapop - 5 GPN = Great Plains Network - 9 LEARN = Lonestar Education and Research Network - 10 MREN = Metropolitan Research and Education Network - 9 NOX = The Northern Crossroads - 22 MAX = Mid-Atlantic Crossroads - 25 Tionesta OADM - 2 Reno OADM - 4 San Francisco OADM - 6 San Luis Obispo OADM - 3 Ogden OADM - 4 Boise OADM - 5 Rieth OADM - 2 Omaha OADM - 3 Tulsa OADM - 10 Jacksonville OADM - 10 Nashville OADM - 4 Birmingham OADM - 5 Buffalo OADM - 3 Albany OADM - 5 New York OADM - 11 Charlotte OADM - 3 Arkansas State University Auburn University Baylor College of Medicine Baylor University Berklee College of Music Binghamton University Boston College Colorado State University Columbia University Complete Genomics Inc. Dartmouth College Desert Research Institute EBSCO Information Services Emory University
----------	--

---

---

Florida Agricultural and Mechanical University  
 Florida Atlantic University  
 Florida International University  
 Florida State University  
 George Mason University  
 George Washington University (The)  
 Georgetown University  
 Georgia Institute of Technology  
 Georgia State University  
 Jet Propulsion Lab  
 Lehigh University  
 Library of Congress  
 Louisiana State University  
 Loyola University Chicago  
 Manhattan School of Music  
 Miami University-Oxford  
 Michigan State University  
 Microsoft Research  
 Mississippi State University  
 Montana State University - Bozeman  
 NASA Goddard Space Flight Center  
 National Archives and Records Administration  
 National Institutes of Health  
 National Library of Medicine  
 National Science Foundation  
 Naval Postgraduate School  
 New Mexico State University -Main Campus  
 New World Symphony  
 New York University  
 Northern Illinois University  
 Northwestern University  
 Oak Ridge National Laboratory  
 Oklahoma State University- Main Campus  
 Old Dominion University  
 Oregon Health & Science University  
 Oregon State University  
 Pepperdine University  
 Portland State University  
 Portland State University  
 Princeton University  
 Purdue University Main Campus  
 Red Hat, Inc.  
 Rice University  
 Rochester Institute of Technology  
 Rutgers, The State University of New Jersey  
 Saint Louis University  
 San Diego State University  
 South Dakota School of Mines and Technology  
 South Dakota State University  
 Southern Illinois University Carbondale  
 Southern Methodist University  
 Stanford University  
 Temple University  
 Texas Christian University  
 Texas Tech University  
 The Children's Hospital of Philadelphia  
 The City University of New York  
 The Ohio State University  
 The Philadelphia Orchestra Association  
 The University of Montana  
 Tufts University  
 Tulane University  
 UCAR/NCAR (University Corporation for Atmospheric Research)  
 United States Antarctic Program (USAP)  
 United States Department of Agriculture, Agricultural Research Services (USDA)  
 United States Department of Commerce Boulder Labs  
 University At Albany, State University of New York  
 University At Buffalo, The State University of New York  
 University of Akron, Main Campus, The  
 University of Alabama at Birmingham  
 University of Alabama in Huntsville  
 University of Alabama, The  
 University of Alaska Fairbanks  
 University of Arizona  
 University of Arkansas at Little Rock

---

---

University of Arkansas for Medical Sciences  
University of Arkansas Main Campus  
University of California, Office of the President  
University of California-Berkeley  
University of California-Davis  
University of California-Irvine  
University of California-Los Angeles  
University of California-Riverside  
University of California-San Diego  
University of California-San Francisco  
University of California-Santa Barbara  
University of California-Santa Cruz  
University of Central Florida  
University of Chicago  
University of Cincinnati Main Campus  
University of Colorado at Boulder  
University of Colorado Denver  
University of Connecticut  
University of Delaware  
University of Denver  
University of Florida  
University of Georgia  
University of Hawaii At Manoa  
University of Houston  
University of Idaho  
University of Illinois at Chicago  
University of Illinois at Urbana-Champaign  
University of Iowa  
University of Kansas Main Campus  
University of Kentucky  
University of Louisville  
University of Maine  
University of Maryland Baltimore County  
University of Maryland College Park  
University of Massachusetts  
University of Memphis  
University of Miami  
University of Michigan-Ann Arbor  
University of Minnesota-Twin Cities  
University of Mississippi  
University of Missouri - Columbia  
University of Missouri - Kansas City  
University of Missouri - Saint Louis  
University of Nebraska - Lincoln  
University of Nevada, Reno  
University of Nevada-Las Vegas  
University of North Texas  
University of Notre Dame  
University of Oklahoma Norman Campus  
University of Oregon  
University of Pittsburgh, Pittsburgh Campus  
University of Puerto Rico  
University of Rhode Island  
University of Rochester  
University of South Carolina - Columbia  
University of Texas at Austin  
University of Texas at El Paso  
University of Texas at San Antonio  
University of Texas Medical Branch At Galveston  
University of Texas Southwestern Medical Center at Dallas  
University of Toledo  
University of Tulsa  
University of Vermont  
Utah State University  
Vanderbilt University  
Virginia Commonwealth University  
West Virginia University  
Wichita State University  
Worcester Polytechnic Institute  
Wright State University  
Yale University  
Yeshiva University  
JSTOR

---

---

CENIC - 16	PNG = Pacific Northwest Gigapop - 10 CENIC = Corporation for Education Network Initiatives in California - 16 UEN = Utah Education Network - 6 Indiana GigaPoP - 8 MissiON = Mississippi Optical Network - 5 MARIA = Mid-Atlantic Research Infrastructure Alliance - 7 Sacramento OADM - 6 Eureka OADM - 2 Sunnyvale OADM - 8 Cornell University DePaul University Missouri University of Science and Technology Morgan State University New Jersey Institute of Technology Northrop Grumman Corporation Rensselaer Polytechnic Institute Syracuse University
------------	---

---

Tabla 4.2: Composición de cada módulo proporcionado por los modelos de bloques estocásticos.

Según la tabla anterior, lo primero que podemos observar es que la estructura comunitaria ofrece un reparto modular desproporcionado, pues el tamaño de los módulos es de 75, 7, 28, 184 y 17 nodos respectivamente. Además, la estructura comunitaria no se ajusta a la partición real de la red. Por ejemplo, el segundo módulo es totalmente desconexo e ignora por completo la disposición geográfica de las universidades que lo forman. Esto se debe al procedimiento en el que se basan los SBM, pues al ser los nodos de cada comunidad estocásticamente equivalentes, a la hora de replicar la red el algoritmo tiende a conectar nodos que originalmente no están asociados. Este problema es habitual en los nodos terminales, puesto que al replicar la red, estos tienen la misma probabilidad de estar conectados con el conector primario o secundario con el que están asociados que con cualquier otro nodo de su módulo.

Por otra parte, como los nodos de los módulos dados por los SBM son estocásticamente equivalentes, el algoritmo no es capaz de detectar la estructura jerárquica de las universidades de Internet2 según estas sean conectores primarios, secundarios o nodos terminales. Si lo hiciera, al replicar la red se generarían conexiones inexistentes entre sus nodos, por lo que esta partición no maximizaría la verosimilitud de los modelos de bloques estocásticos.

A modo de comparativa entre los resultados obtenidos con los algoritmos de Lovaina y SBM podemos afirmar que en las redes tecnológicas de la forma centro-periferia, como la de Internet2, el algoritmo de Lovaina es más efectivo que el de SBM, pues aunque ninguno detecta los ciclos que se forman a lo largo de la red, el primero distingue la disposición geográfica de las universidades. De esta forma, el algoritmo de Lovaina aporta más información de la disposición de este tipo de redes que los SBM, que no distinguen ni la situación geográfica ni la jerárquica de las distintas instituciones de la red.

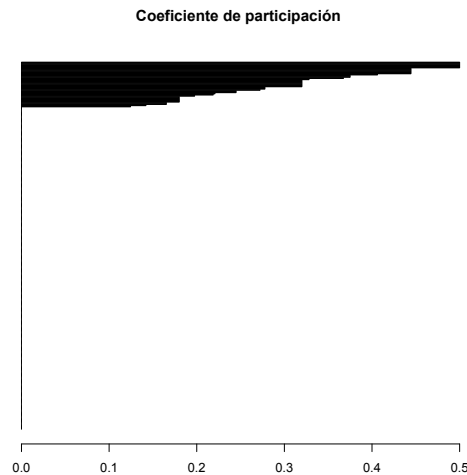


Figura 4.2: Coeficiente de participación de los nodos de la red de Internet2. La disposición de sus valores es progresiva, de menor a mayor valor.

### 4.1.3. Centro y núcleo

Por último, al igual que hicimos con *C. elegans*, realizamos un estudio pormenorizado de las características de los nodos de Internet2 y tratamos de detectar un núcleo funcional. Para ello comenzamos estudiando el coeficiente de participación de la red. Los resultados obtenidos están representados en la Figura 4.2, donde disponemos en orden creciente los valores del coeficiente de participación de la red. Observando detenidamente la imagen podemos comprobar que gran parte de los valores obtenidos son nulos, lo que se asocia a la gran cantidad de nodos terminales presentes en la red que, como vimos en el Capítulo 2, tienen un coeficiente de participación nulo. De esa forma, los 251 nodos terminales de la red tienen un coeficiente de participación nulo y tan solo los 60 restantes son los encargados de conectar las comunidades, permitiendo el intercambio de información en la red.

Además, estudiando los valores concretos, vemos que la red cuenta con 21 nodos conectores, que forman parte del conjunto de ciclos centrales, que no coinciden con los “conectores principales” establecidos por el consorcio, y 290 provinciales. En la Figura 4.3 indicamos los nodos conectores de la red marcándolos por medio de una estrella. Al conectarlos mediante caminos geodésicos obtenemos la red de conectores de Internet2 representada en la Figura 4.4, que cubre la práctica totalidad de los conectores primarios y secundarios dispuestos por el consorcio.

Los valores de la centralidad están representados en la Figura 4.5 dispuestos en orden



Figura 4.3: Representación de los nodos conectores de Internet2.

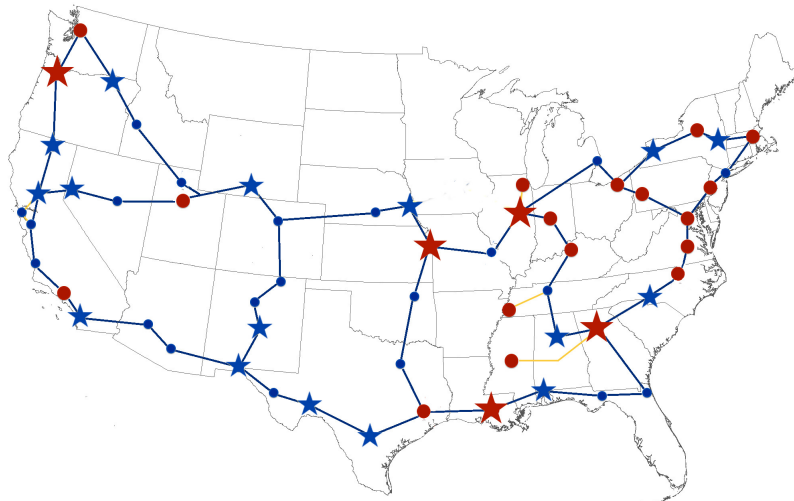


Figura 4.4: Representación de la red de conectores de Internet2.

creciente. Solo las 60 universidades del centro geográfico de la red tienen centralidad no nula ya que los restantes nodos son terminales.

Si definimos el centro como los nodos con coeficiente de centralidad superior a 0.1 obtenemos que este está conformado por 39 nodos. Procede comprobar si este conjunto de universidades podría actuar como un núcleo funcional de la red. Para ello estudiamos la composición resultante de eliminar dichas universidades del consorcio. Al hacerlo la red se disocia en 8 componentes conexas de 19, 13, 11, 9, 5, 4, 3 y 2 nodos y 206 universidades aisladas. Esto indica que si este centro dejara de transmitir información, la red pasaría a

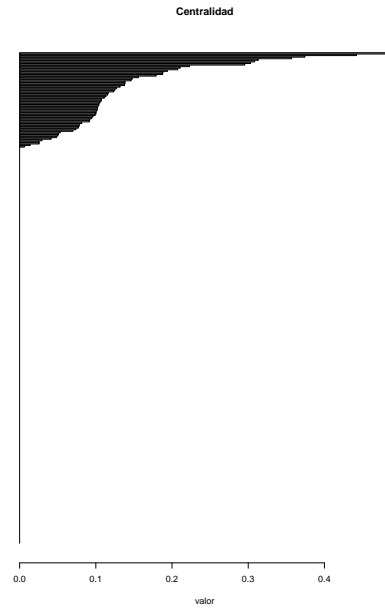


Figura 4.5: Coeficiente de centralidad de los nodos de la red de Internet2. La disposición de sus valores es progresiva, de menor a mayor valor.

ser disconexa y solo en algunos casos habría una colaboración a pequeña escala. Esto nos sugiere que este es un buen candidato a núcleo funcional.

Por otro lado, si cambiamos el criterio y tomamos como límite para el coeficiente de centralidad el promedio de los valores presentes en la red entonces tenemos 56 universidades centrales. Como es de esperar, todos estos centros de enseñanza forman parte del centro geográfico de la red. Si estudiamos de nuevo la composición de la red al eliminar este conjunto, vemos que queda dividida en 3 componentes conexas distintas, dos de ellas de 5 nodos y la otra de 2, y 243 universidades aisladas. De esa forma, si este conjunto de nodos no estuviera presente, la información dejaría de fluir por la red, por lo que este centro de la red sí que actúa como un núcleo funcional de la misma. En la Figura 4.6 representamos la disposición de este subconjunto en el núcleo central.

Si combinamos los criterios de participación y centralidad, considerando el conjunto de nodos que son al mismo tiempo conectores y centrales, tendríamos 21 nodos. Al suprimir estos nodos, quedan 14 componentes conexas, sin contar las universidades que quedarían aisladas, pero la componente conexa más grande tiene un tamaño de 100 nodos, lo que representa un 32,15% de la red. Por tanto no podemos afirmar que este sea un núcleo funcional. En la Figura 4.7 representamos el núcleo central tras suprimir este subconjunto de la red.

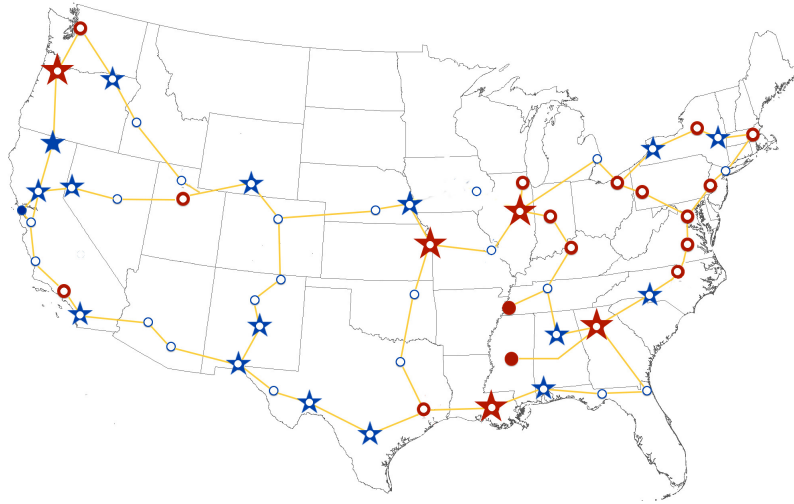


Figura 4.6: Representación de los nodos centrales de Internet2 con el criterio del promedio.

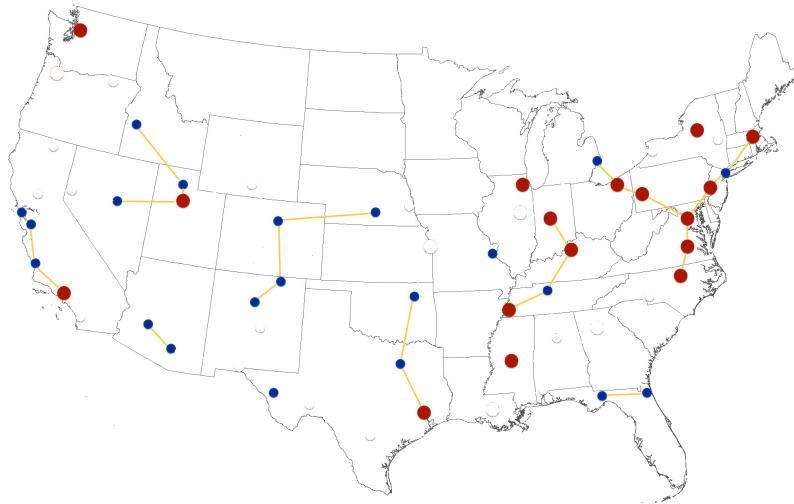


Figura 4.7: Núcleo central de la red tras suprimir el conjunto de nodos que son al mismo tiempo conectores y centrales.

Ahora bien, si consideramos la red de conectores descrita en la Figura 4.4 y que está formada por 54 universidades, vemos que su eliminación nos deja 5 componentes conexas de 21, 9, 5, 5 y 2 nodos y 215 universidades aisladas. Por tanto, también puede ser considerado como un núcleo funcional de la red Internet2.

Por todo ello, podemos afirmar que el algoritmo de Lovaina es capaz de detectar la estructura comunitaria de la red telemática Internet2 de modo bastante aceptable, aunque no sea capaz de detectar los ciclos que forman la estructura central de la red. Además, tal y como pasaba en *C. elegans*, el algoritmo proporciona por medio de la red de conectores un

núcleo funcional razonable, si bien otro núcleo muy similar puede ser descrito en términos de centralidad.

## 4.2. Grafo jerárquico de Ravasz

Por último queremos saber cómo de efectivos son los algoritmos para la detección de la estructura comunitaria en las redes jerárquicas. Para ello empleamos el grafo jerárquico de Ravasz, que fue previamente introducido y cuya estructura comunitaria viene dada por sus cuatro bloques: el central y los tres periféricos.

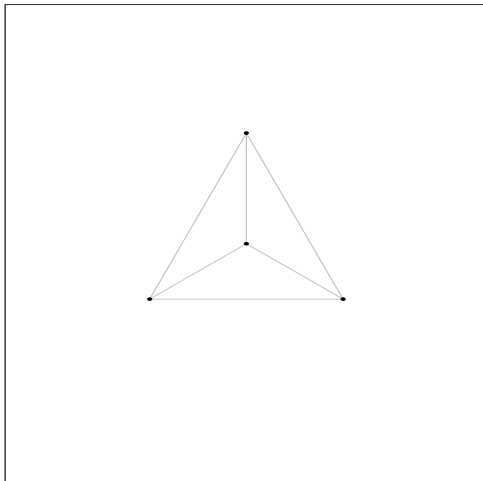
### 4.2.1. Algoritmo de Lovaina

Comenzamos trabajando con el algoritmo de Lovaina. Aplicamos este método a las redes jerárquicas de grado 1, 2, 3 y 4, y podemos comprobar inductivamente si es capaz de distinguir los módulos constructivos.

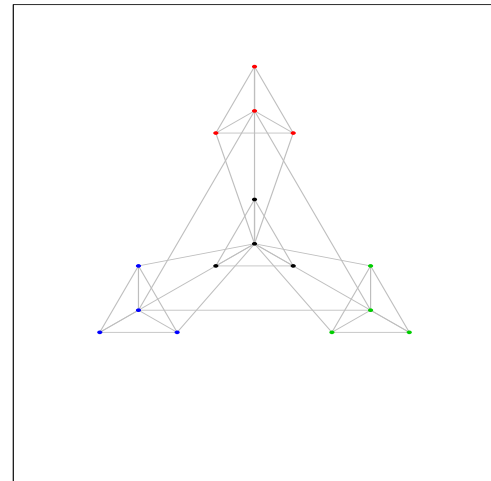
Así vemos que la modularidad de los grafos jerárquicos de orden 1, 2, 3 y 4 toma valores 0, 0.414, 0.534 y 0.606 respectivamente. El valor de modularidad nulo asociado al grafo jerárquico de primer orden se debe a que este es un grafo completo con 4 nodos, por lo que todas las aristas representadas son las esperadas en un diagrama equivalente. Además, conforme el grafo se va diferenciando más del grafo completo la modularidad va aumentando.

Por otra parte, por lo que se refiere a la estructura comunitaria, esta se describe en la Figura 4.8, donde vemos que el algoritmo de Lovaina es capaz de detectar la estructura comunitaria esperada en orden 1, 2 y 3. Sin embargo, para el grafo jerárquico de orden 4, la estructura comunitaria propuesta no se corresponde con la estructura real, sino que el algoritmo detecta 10 comunidades diferentes, seis con tamaño 16, dos de 48, una de 36 nodos y otra de 28. Claramente no parece una partición aceptable, pues el algoritmo no detecta el bloque central y los bloques periféricos, sino que los separa en varias subcomunidades y deshace la simetría del grafo agrupando nodos del bloque central con vértices ubicados en una de las periferias.

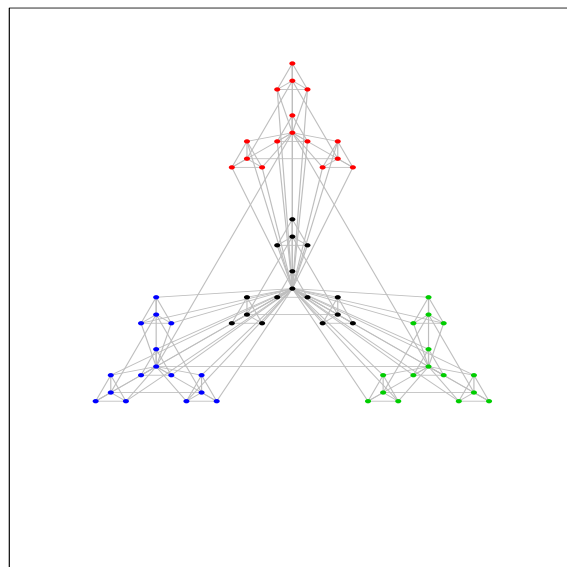
Además, otra prueba de que el algoritmo de Lovaina no funciona correctamente en la red jerárquica de orden 4 es que, a pesar ser un grafo simétrico, el reparto de los nodos en los bloques periféricos no es idéntico en ningún caso. Esto sugiere que a medida que aumenta la complejidad de la red jerárquica, el algoritmo va obteniendo resultados más desacordes a la distribución real.



(a) Grafo jerárquico de orden 1.

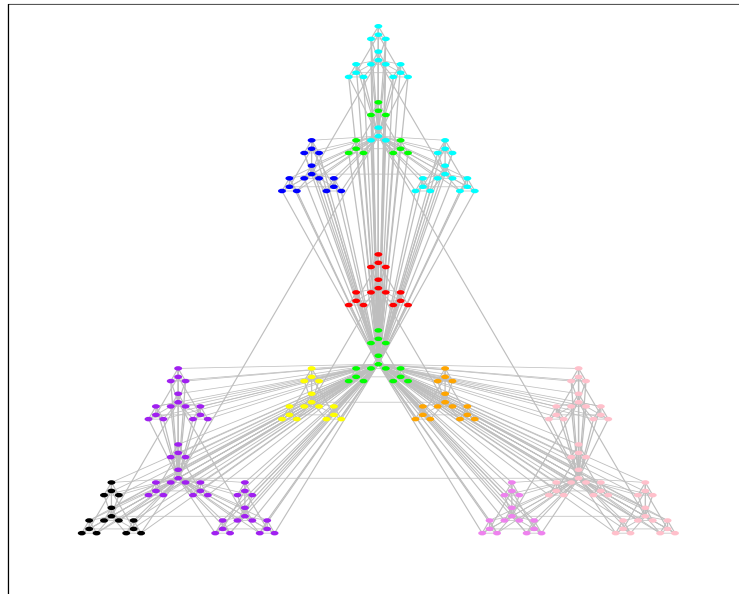


(b) Grafo jerárquico de orden 2.



(c) Grafo jerárquico de orden 3.

Figura 4.8: Particiones con el algoritmo de Lovaina del grafo jerárquico de Ravasz.



(d) Grafo jerárquico de orden 4.

Figura 4.8: Particiones con el algoritmo de Lovaina del grafo jerárquico de Ravasz.

#### 4.2.2. Modelos de bloques estocásticos

Por otra parte, como en los casos anteriores, estudiamos también si los modelos de bloques estocásticos son capaces de detectar la estructura comunitaria del grafo de Ravasz de órdenes 1, 2, 3 y 4.

Como dijimos al aplicar el algoritmo de Lovaina, sabemos que la estructura comunitaria del grafo jerárquico viene dada por cuatro módulos que se relacionan con los 4 bloques principales del grafo: el central y los tres periféricos. Por ello, aplicamos los modelos de bloques estocásticos, pidiéndole al algoritmo que nos devuelva cuatro módulos. De esa forma, mediante un proceso de inferencia y con la aplicación de métodos estadísticos y procesos de máxima verosimilitud, obtenemos los resultados presentados en la Figura 4.9.

Podemos ver que los modelos de bloques estocásticos nos devuelven cuatro comunidades, aunque tampoco son capaces de detectar la estructura comunitaria del grafo jerárquico. Sin embargo, los resultados obtenidos sobre esta red merecen una especial atención.

Los modelos de bloques estocásticos, como dijimos, tienen la peculiaridad de que buscan bloques cuyos nodos cuenten con una probabilidad de conexión similar con nodos de otros bloques, es decir, nodos que tengan patrones de conectividad similares. Esto motiva que proponga una distribución modular interesante, pues el algoritmo es capaz de distinguir

entre centro y periferia, al igual que pasaba con el grafo tipo estrella comentado en el Capítulo 2. Veámoslo detenidamente en cada orden.

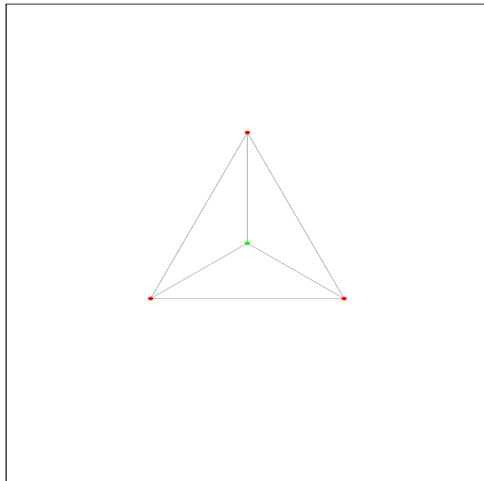
En la Figura 4.9 (a), podemos observar esta división en centro y periferia, aunque al ser este grafo completo, los modelos de bloques estocásticos podrían dar cualquier división, pues la probabilidad de conexión entre los nodos de los distintos bloques, así como de obtener conexiones dentro de cada uno de ellos, es siempre 1. Por tanto, el procedimiento de inferencia al aplicar este método podría dar cualquier resultado.

En la Figura 4.9 (b), relativa al grafo jerárquico de orden 2, podemos observar que los modelos de bloques estocásticos distinguen ligeramente centros y periferias. En él los SBM agrupan en una sola comunidad dos de los módulos periféricos, mientras que en el central y en la otra comunidad de la periferia sí que separa los nodos centrales de los periféricos en dos comunidades. Al estudiar el grafo jerárquico de orden 3 en la Figura 4.9 (c), vemos que este efecto se va remarcando, pues en este caso tan solo es un módulo teórico de la periferia el que SBM agrupa en una comunidad, mientras que en el resto de bloques el algoritmo diferencia el centro y los dos niveles de periferia de la red. Finalmente, cuando el grafo jerárquico es de orden 4 en la Figura 4.9 (d), los niveles jerárquicos se distinguen completamente y el algoritmo separa centros y periferias. Al aumentar la complejidad del grafo, haciendo que este se parezca cada vez menos a un grafo completo, el algoritmo diferencia con mayor facilidad la disposición de los nodos en la red jerárquica.

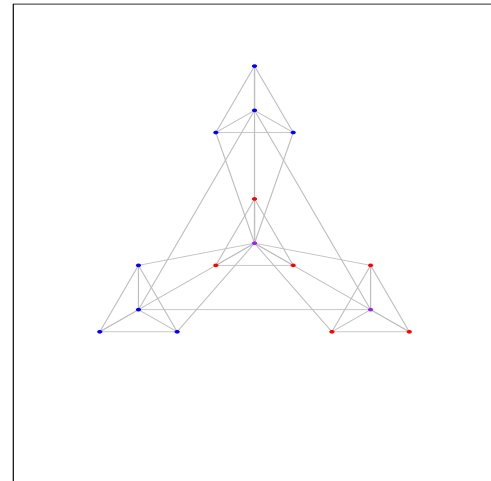
En el grafo jerárquico de orden 4 podemos ver que los módulos se toman de forma que maximizan la probabilidad de conexión entre nodos de cada bloque. Así, la comunidad dada por los centros de los bloques periféricos tienen una probabilidad de conexión interna total, nula con la comunidad dada por el nodo central de la red, algo elevada con la comunidad dada por los nodos del módulo azul y mucho más alta con los vértices del módulo verde. Es el modo en que la máxima verosimilitud y el procedimiento inferencial generan la red y determina su partición en módulos.

Por este motivo, conjeturamos que si la complejidad de la red continuara aumentando, el algoritmo seguiría separando los centros de diversos órdenes de las periferias. Además, por la simetría del grafo, a los nodos que ocupan posiciones simétricas se les puede dotar de una función idéntica en la red. Los SBM detectan esta funcionalidad en sus comunidades, agrupando aquellos nodos que tienen una situación jerárquica equivalente.

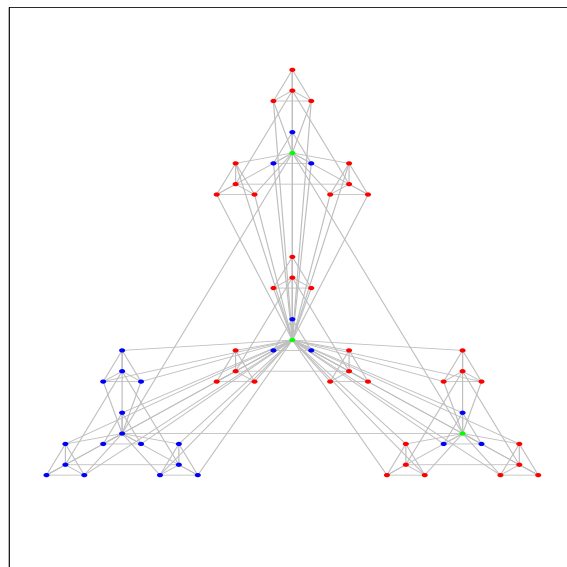
Como comentario final, podemos afirmar que ni el algoritmo de Lovaina ni los modelos estocásticos son capaces de detectar la estructura comunitaria para cualquier orden del grafo. No obstante, los modelos de bloques estocásticos son útiles para hallar patrones de construcción de redes modulares jerárquicas.



(a) Grafo jerárquico de orden 1.

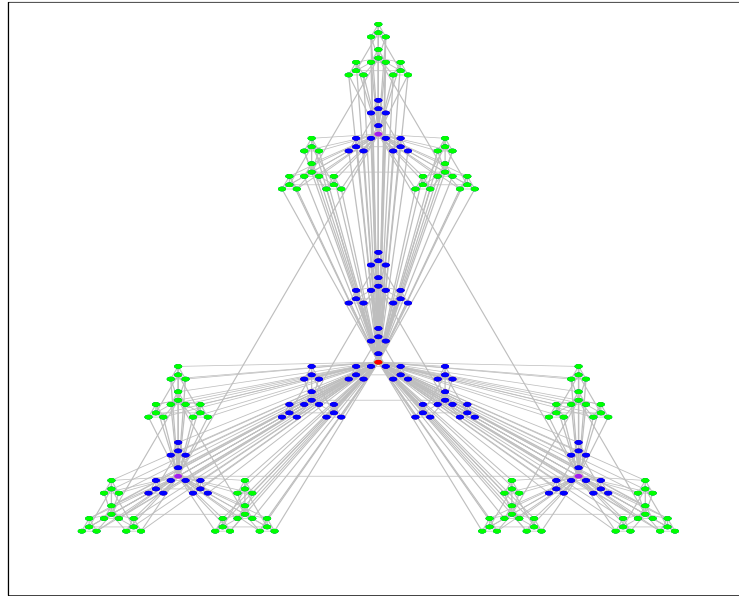


(b) Grafo jerárquico de orden 2.



(c) Grafo jerárquico de orden 3.

Figura 4.9: Estructura comunitaria propuesta por el algoritmo SBM para los grafos jerárquicos de los tres primeros órdenes.



(d) Grafo jerárquico de orden 4.

Figura 4.9: Estructura comunitaria propuesta por el algoritmo SBM para el grafo jerárquico de orden 4.

### 4.2.3. Centro y núcleo

Finalmente, también estudiamos los coeficientes de participación y de centralidad de los 256 nodos de la red jerárquica de Ravasz y tratamos de determinar su núcleo funcional. Comenzamos estudiando el coeficiente de participación.

La Figura 4.10 muestra en orden creciente y dispuestos de forma escalonada los coeficientes de los nodos. Esto se debe a la simetría del grafo y de parte de los módulos, lo que motiva que nodos en situaciones idénticas tengan los mismos valores de participación. Así vemos que hay 130 nodos conectores y 126 provinciales. Entre los primeros cabe destacar la presencia del nodo central de la red y de los nodos centrales de las periferias, ya que estos nodos tienen el coeficiente de participación más elevado. Además, los nodos conectores forman un subgrafo conexo. Esto implica que la red de conectores coincide exactamente con el conjunto de nodos conectores.

Por otro lado, de acuerdo con el coeficiente de centralidad, esperaríamos que los cuatro nodos ubicados en la posición central de los bloques principales sean centrales. Al aplicar el coeficiente de centralidad obtenemos los resultados representados en la Figura 4.11. En ella vemos que hay un único nodo con un coeficiente de centralidad superior a 0.5. Este

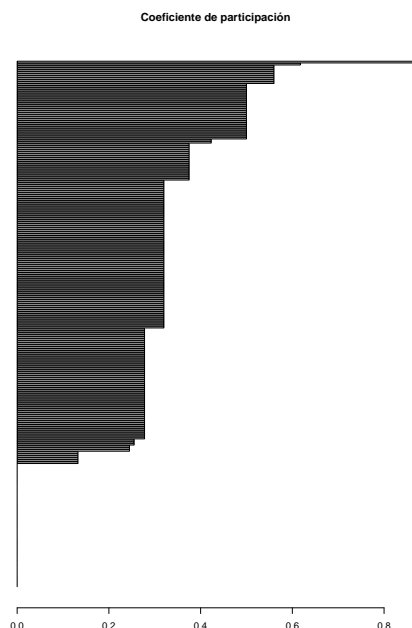


Figura 4.10: Representación del coeficiente de participación de los nodos de la red de Ravasz de orden 4 dispuesto en orden creciente.

nodo es el centro geométrico del grafo, que se encarga de conectar los nodos periféricos de distintos niveles del grafo. Para transmitir información entre dos nodos periféricos, en general, es conveniente pasar por el centro, pues no hay un camino de menor distancia que los conecte. A continuación, podemos ver que los siguientes nodos según el valor del coeficiente de centralidad son los tres centros de los bloques periféricos. La interpretación del valor de su coeficiente de centralidad en comparación al del resto de la red es la misma que la del centro global, solo que estos nodos actúan como puente dentro de su bloque.

Si definimos el centro como el conjunto de nodos cuya centralidad supera el valor de 0.1, entonces tenemos que este estaría formado únicamente por el nodo central de la red. Este nodo no define un núcleo funcional de la red, pues una de las 4 componentes que resultan al eliminarlo tiene 192 nodos. Por otra parte, tomando como límite del coeficiente de centralidad el promedio de los valores de este en los nodos de la red, incorporamos a los cuatro nodos centrales. En este caso, si los suprimimos, la red se desconecta en 12 componentes conexas, donde las de mayor tamaño cuentan con 48 vértices.

Por último, estudiamos si los nodos conectores centrales o la red de conectores definen un núcleo funcional en la red jerárquica de Ravasz. Comenzamos estudiando el primer conjunto. Para ello consideramos centrales los nodos que superen el valor medio del coeficiente de

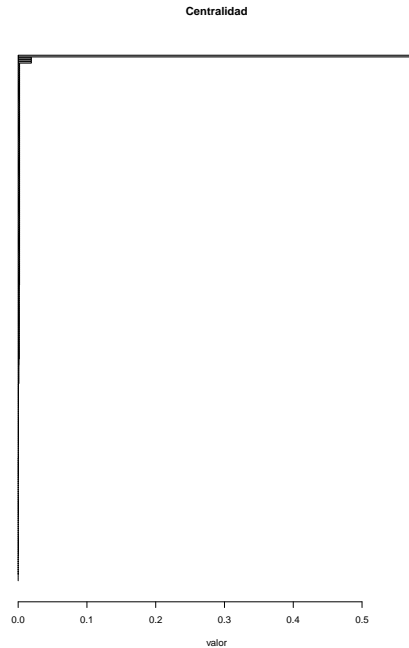


Figura 4.11: Representación del coeficiente de centralidad de los nodos de la red jerárquica de cuarto orden dispuesto en orden creciente.

centralidad en la red. Al calcular el conjunto obtenemos que este coincide precisamente con el centro de la red, por lo que no podemos afirmar que los nodos conectores centrales constituyan un núcleo funcional del grafo. Como comentario adicional cabe destacar que este conjunto es precisamente el *rich club* de la red, es decir, el conjunto de nodos de mayor grado del grafo. Al igual que pasaba en el conectoma de *C. elegans*, obtenemos que a pesar de que este conjunto constituye un grupo de nodos de gran importancia para el funcionamiento de la red, este no es imprescindible, pues sin él la red todavía mantendría cierta funcionalidad a nivel local.

Por otra parte, al considerar la red de conectores, que como sabemos coincide con el conjunto de nodos conectores de la red, obtenemos las componentes conexas del grafo complementario representadas en la Figura 4.12. Podemos ver que hay 15 componentes conexas, donde 3 cuentan con 27 nodos cada una. Por tanto, no está claro que este subgrafo constituya un núcleo funcional de la red, pues pese a que la red se disocia en pequeñas componentes, de los 256 nodos originales todavía hay 81 que mantienen conectividad local.

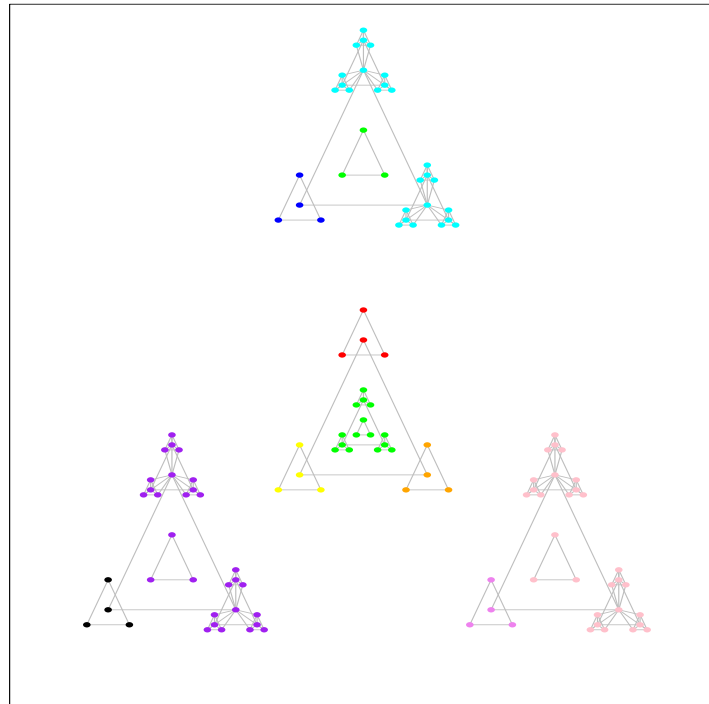


Figura 4.12: Representación del grafo sin la red de conectores.



## Capítulo 5

# Comparación de métodos

En este capítulo comparamos los resultados obtenidos en las redes estudiadas con cada algoritmo y vemos en qué medida las redes de control nos indican la capacidad de los algoritmos de detectar la estructura comunitaria en el conectoma de *C. elegans*.

Además, estudiamos qué tipo de estructura tienen las tres redes propuestas en base a su matriz de conectividad y vemos su posible dependencia de la ordenación tomada de los nodos de la red.

### 5.1. Algoritmo de Lovaina

Comenzamos estudiando los resultados obtenidos tras aplicar el algoritmo de Lovaina en las tres redes tratadas en este trabajo.

Como dijimos en el Capítulo 2, el algoritmo de Lovaina trata de detectar la estructura comunitaria de las redes mediante la búsqueda de particiones con una alta densidad de conexiones en cada módulo. De esa forma, el algoritmo no atiende a propiedades geométricas de las redes, lo que explica que en redes construidas alrededor de un núcleo cíclico, como es la red asociada al consorcio Internet2, el algoritmo de Lovaina no sea capaz de detectar ese núcleo, sino que agrupa las comunidades en torno a pequeños bloques centrales. Por otra parte, en redes jerárquicas como la de Ravasz, el algoritmo de Lovaina no es capaz de detectar los patrones repetitivos que generan este tipo de redes. También se aprecia la incapacidad del algoritmo de Lovaina para detectar la simetría del grafo cuando su complejidad es suficientemente alta para diferenciarlo de un grafo completo.

Esto puede justificar el hecho de que al aplicarle el algoritmo de Lovaina al conectoma de *C. elegans*, donde la modularidad es de 0.414, no exista una relación clara entre módulos y ganglios, salvo en el caso de la faringe. Esto se debe a que los ganglios agrupan las neuronas por razones espaciales y no necesariamente por razones de conectividad o funcionales. Sin

embargo, debido a la particular geometría del conectoma de *C. elegans* y a la disposición de los ganglios a lo largo de su cuerpo, no podemos asegurar que la  $Q$ -modularidad de otras redes cerebrales más complejas sea más alta y aumente la relación entre módulos y ganglios.

Los resultados relativamente bajos de la modularidad de las redes cerebrales humanas obtenidas mediante imagen por resonancia magnética fMRI y DSI [9] y mostradas en [3], con valores de 0.24 y 0.56 respectivamente, están justificadas por tratarse de conectomas a media escala, que no están formados por neuronas, sino por regiones de interés (104 en el primer caso y 998 en el segundo). Los resultados son aún menores en [10], aunque se debe a las limitaciones de los primeros algoritmos usados para determinar la modularidad.

A pesar de esto, el hecho de que el algoritmo de Lovaina detecte conjuntos de nodos densamente interconectados tiene una parte positiva, ya que permite detectar subconjuntos que ejecutan funciones conjuntas a media escala. Esto lo podemos observar mediante las redes de control: en Internet2, agrupa las instituciones académicas en pequeños grupos regionales, obteniendo una modularidad de 0.866 y una división en 18 módulos; en el grafo jerárquico de Ravasz, aunque no conserva la simetría del grafo, el algoritmo es capaz de mantener la agrupación en submódulos regionales, obteniendo una modularidad de 0.606. Este mismo hecho ya estaba presente en estudios previos, pero con resultados ligeramente diferentes: en [1] la modularidad de Internet2 y Ravasz era de 0.86 y 0,61 respectivamente, mientras que en [14], para Internet2 se obtuvo una modularidad de 0.859 y una división de la red en 16 módulos. Por tanto, comparando estos resultados con los obtenidos para el conectoma de *C. elegans* podemos afirmar que, aunque el algoritmo de Lovaina no es capaz de detectar los ganglios del conectoma, muestra una partición en módulos altamente interconectados con una cierta disposición espacial y en algunos casos formados por submódulos regionales vinculados con los ganglios.

Por último, en cuanto a la detección de un núcleo funcional en las redes, procede señalar que cuanto mayor sea la complejidad de la red, mayor tendrá que ser el tamaño relativo de su núcleo.

En la red Internet2, detectamos dos núcleos similares, uno obtenido mediante el conjunto de nodos centrales, tomando como límite la media de la centralidad, y otro a través de la red de conectores, de 56 y 54 universidades respectivamente. Además, la simplicidad de la red, que cuenta con 251 nodos terminales, permite que al considerar dichos conjuntos como núcleos funcionales obtengamos que estos son imprescindibles para que la red funcione correctamente, tanto a nivel global como localmente, pues eliminarlos supondría aislar a una gran parte de las universidades que actúan como nodos terminales.

En el conectoma de *C. elegans*, la complejidad de la red y la gran cantidad de sinapsis de control hacen inviable la definición de un núcleo funcional en la red con un tamaño

moderado. Este hecho provoca que conjuntos previamente estudiados en la bibliografía como el *housglass weit* o cuello de botella y el *rich club* que son conjuntos de neuronas importantes en la transmisión de información a lo largo de la red, no sean imprescindibles para el funcionamiento de la red a media escala. Tan solo la red de conectores proporciona un núcleo funcional pero está formado por el 84,67% de la red original. En contra de lo que sugieren algunos estudios, no parece que sea posible definir un pequeño núcleo funcional en la red neuronal de un ser vivo, ya que por razones evolutivas debe mantener una parte esencial de sus funciones pese a las posibles mutaciones o al ataque de patógenos.

## 5.2. Modelos de bloques estocásticos

Por otra parte, repasamos los resultados obtenidos con los modelos de bloques estocásticos, viendo hasta qué punto detectan la estructura comunitaria del conectoma de *C. elegans* y su utilidad en el estudio de las redes neuronales de otros seres vivos.

Como vimos, el algoritmo SBM busca la similitud de los patrones de conectividad de los nodos para agruparlos según ese criterio.

Aplicando el algoritmo a las redes de control obtuvimos resultados dispares, pues mientras que el algoritmo no detectaba una estructura comunitaria aceptable en Internet2, pues no identificaba ni los ciclos ni la posición de los nodos centrales, el algoritmo era capaz de detectar la posición relativa de los nodos, según pertenecieran al centro o a los distintos niveles periféricos, en la red jerárquica de Ravasz. Esto se debe a que la construcción de las redes es distinta en cada caso, lo que da lugar a esta diferencia en los resultados obtenidos. En efecto, la red jerárquica de Ravasz se construye mediante un proceso iterativo, lo que propicia que a lo largo de la red aparezcan nodos que ocupan posiciones simétricas en el grafo, dotando a estos nodos de una funcionalidad idéntica en la red. Esta relación entre los nodos de la red es la que detectan los SBM. Por otra parte, en Internet2 no existe tal funcionalidad, pues la red fue construída con el objetivo de conectar las instituciones académicas según su situación geográfica, sin seguir un motivo práctico en particular.

Por su parte, en el caso del conectoma de *C. elegans*, como veíamos en la Figura 3.6, el algoritmo era capaz de privilegiar determinadas funciones en los módulos del modelo. Esto puede deberse a que neuronas encargadas de realizar funciones similares tienen a menudo patrones de conectividad parecidos, por lo que las neuronas, ya sean sensoriales, interneuronas, motoras o de la faringe, tienden a agruparse según su función. En efecto, como las neuronas sensoriales están encargadas de registrar los estímulos externos, las interneuronas de transmitir los mensajes producidos por las neuronas sensoriales a las motoras y estas últimas de responder a los estímulos, es razonable pensar que los patrones

de conectividad son similares en términos estadísticos.

De hecho, hay un uso creciente de los modelos de bloques estocásticos en el estudio de las redes cerebrales de diferentes especies. Los trabajos de Pavlovic et al. [13] y más recientemente Betzel et al. [22] son una buena muestra. No obstante, salvo el conectoma de *C. elegans*, todas esas redes cerebrales están constituidas por regiones de interés con arquitecturas diferentes de los auténticos conectomas.

No obstante, puede esperarse que los modelos de bloques estocásticos agrupen las neuronas según criterios funcionales, pues neuronas de un mismo tipo deben seguir patrones similares, siendo necesario en general que los estímulos captados por las neuronas sensoriales pasen por las interneuronas para llegar a las motoras, siguiendo la información un esquema similar en especies similares a *C. elegans*.

### 5.3. Tipo de estructura y dependencia del algoritmo

Por último estudiamos el tipo de estructura (asortativa, núcleo-periferia o disasortativa) de las tres redes analizadas a lo largo del trabajo. Para ello examinamos la matriz de adyacencia de las redes, tratando de detectar su particular disposición.

Comenzamos estudiando el caso de *C. elegans*. En primer lugar discutimos el tipo de estructura del conectoma mediante la matriz de adyacencia de la red según el orden original de [20] que se corresponde con el reparto funcional. Su representación en la Figura 5.1, muestra la división de las neuronas en cuatro grupos: faringe, neuronas sensoriales, interneuronas y neuronas motoras. En la matriz observamos tres grandes bloques con una estructura vagamente asortativa, donde el primero se corresponde con las neuronas de la faringe, el segundo con un bloque de neuronas sensoriales, interneuronas y una parte de motoras y un tercer bloque formado por las restantes neuronas encargadas de la motricidad de *C. elegans*.

Destaca el caso de la faringe, que tan solo cuenta con 4 sinapsis externas, lo que propicia que los algoritmos de detección la identifiquen como un módulo particular con claridad. Los otros dos bloques son más difusos, pues la densidad de conexiones no es tan elevada como en la faringe y aparecen ligeros patrones disasortativos.

Las matrices de adyacencia obtenidas al ordenar las neuronas a partir de la estructura comunitaria que proporcionan los algoritmos de Lovaina y SBM pueden verse en la Figura 5.2.

Los cinco módulos dados por el algoritmo de Lovaina parecen determinar una estructura asortativa. Sin embargo, la matriz muestra ciertas componentes disasortativas debidas a la densidad de conexiones entre la segunda y tercera comunidad y entre la cuarta y quinta

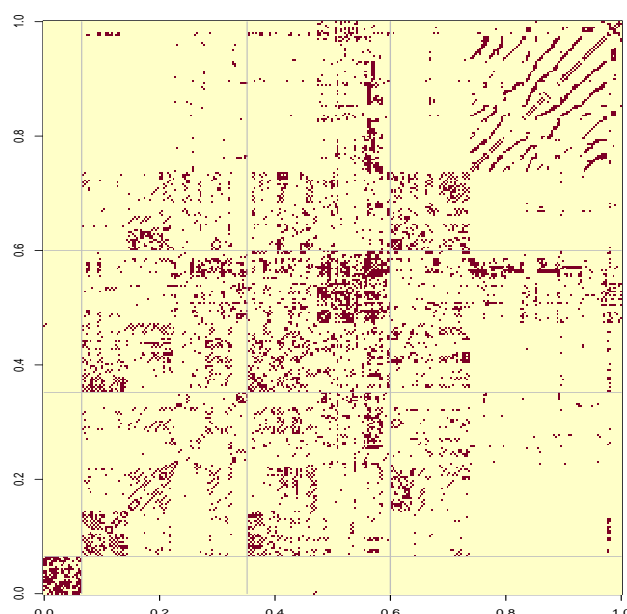


Figura 5.1: Representación de la matriz de adyacencia de *C. elegans* según su reparto funcional.

comunidad, que parece obedecer tanto a razones anatómicas (véase la Figura 3.1) como funcionales (véase la Figura 3.3).

Por último, en la matriz de adyacencia de *C. elegans* ordenada siguiendo el reparto por módulos de SBM vemos que los bloques correspondientes al segundo, tercer y cuarto módulo tienen una estructura asortativa, pero el quinto módulo, formado por interneuronas, mantiene una alta conectividad con los restantes módulos, excepto con el segundo, que corresponde a la faringe.

Aunque las matrices de adyacencia del conectoma de *C. elegans* mantienen ciertas características asortativas con independencia del orden que proporciona cada tipo de reparto, funcional, modular o por bloques estocásticos, la conclusión a la que llegamos al comparar las tres matrices es que el tipo de estructura depende del método usado para determinar la estructura comunitaria. Luego las afirmaciones sobre los tipos de estructura de las redes, habituales en análisis de redes y neurociencia, no están justificadas desde el momento en el que dependen de los métodos de determinación de la estructura comunitaria.

En segundo lugar estudiamos la naturaleza de la matriz de adyacencia de la red Internet2. Para ello ordenamos los nodos según la base de datos del consorcio con un bloque inicial de conectores primarios y secundarios seguido de una lista de universidades conectadas a esos nodos principales. En la Figura 5.3 podemos ver la matriz correspondiente. En ella se distingue un conjunto de nodos conectados entre sí, que corresponden con conectores primarios y

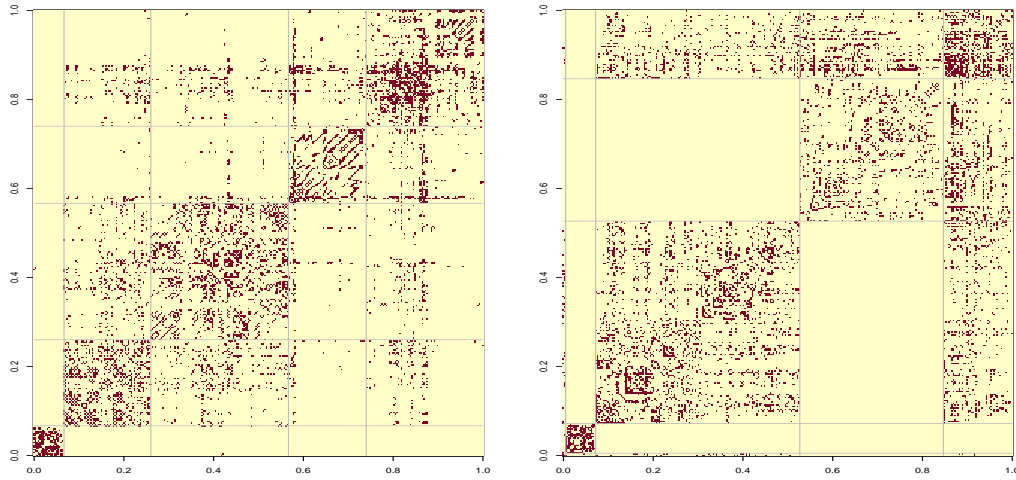


Figura 5.2: Matrices de adyacencia del conectoma de *C. elegans*. A la izquierda usando el reparto modular dado por el algoritmo de Lovaina y a la derecha por el algoritmo SBM.

secundarios que conforman el núcleo de la red, y un conjunto periférico correspondiente a los nodos terminales. Está claro que la red tiene una estructura de tipo centro-periferia.

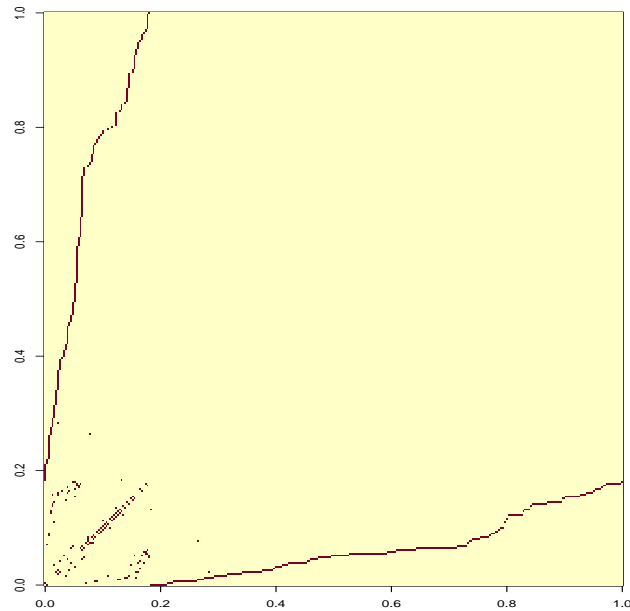


Figura 5.3: Representación de la matriz de adyacencia de Internet2 según su estructura jerárquica.

No obstante, si usamos el reparto modular proporcionado por los algoritmos de Lovaina y SBM, las matrices de adyacencia son diferentes como puede verse en la Figura 5.4.

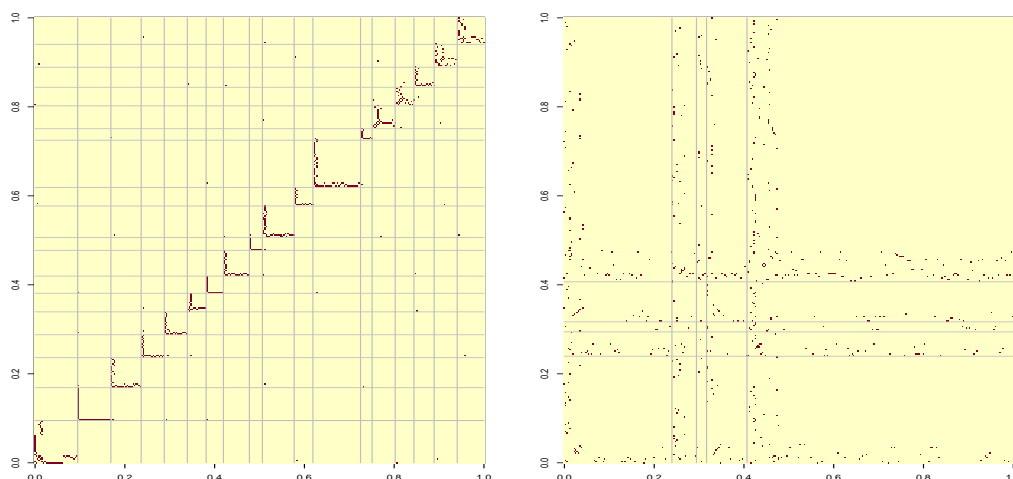


Figura 5.4: Matrices de adyacencia de la red Internet2. A la izquierda según el reparto modular dado por el algoritmo de Lovaina y a la derecha por SBM.

En el primer caso, la matriz de adyacencia muestra una estructura asortativa en la que cada uno de los 18 módulos posee una estructura interna de tipo centro-periferia.

Por el contrario, si usamos el reparto modular obtenido al aplicar el algoritmo SBM, vemos que la matriz de adyacencia tiene una estructura disasortativa.

En este caso resulta aún más claro que la estructura atribuida a la red depende del método usado para ordenar los nodos, lo que impide una atribución correcta, incluso cuando la red posee un tipo de estructura bien definido como ocurre con Internet2. La afirmación inicial de que la red Internet2 es de tipo núcleo-periferia es correcta, pero lo es porque hemos usado el mismo orden que ha empleado el consorcio de universidades para construir la red telemática.

Por último, estudiamos el grafo jerárquico de Ravasz. En primer lugar, igual que en el caso anterior representamos la matriz de adyacencia siguiendo el criterio usado para construirla. En la Figura 5.5 representamos la matriz de adyacencia del grafo de Ravasz de orden 4. En ella podemos observar una estructura asortativa, donde los nodos de los distintos bloques apenas se relacionan entre sí y tan solo el nodo central de la red está conectando los cuatro bloques del grafo. Conviene destacar que cada bloque está dotado de

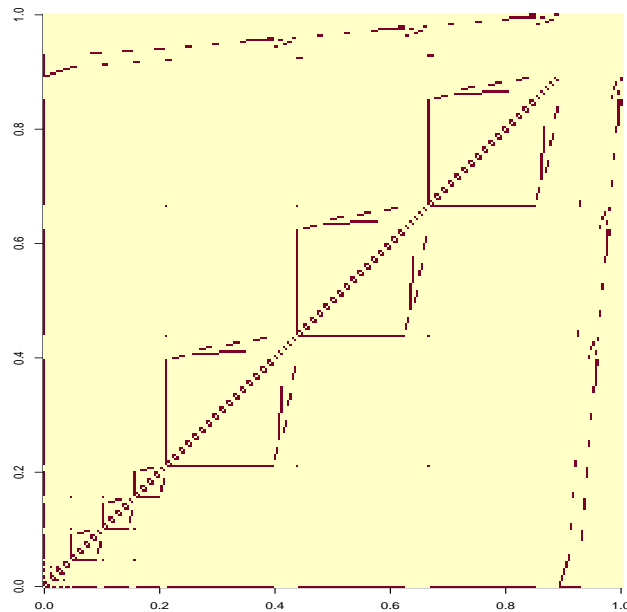


Figura 5.5: Representación de la matriz de adyacencia del grafo jerárquico de Ravasz en base a su construcción.

una estructura local de tipo centro-periferia alrededor de un nodo central.

En la Figura 5.6 presentamos las matrices de adyacencia obtenidas al ordenar los nodos según el reparto modular obtenido mediante los algoritmos de Lovaina y SBM.

En el primer caso, la matriz sigue mostrando una estructura vagamente asortativa, donde las comunidades están densamente conectadas internamente y los nodos de cada módulo cuentan con una baja cantidad de conexiones con vértices de otras comunidades. No obstante, la matriz de adyacencia nos permite confirmar la existencia de un nodo central que conecta toda la red y de tres nodos centrales correspondientes a cada uno de los tres bloques periféricos, encargados de conectar los nodos de sus respectivos bloques.

Por el contrario, la matriz de adyacencia asociada al reparto por bloques estocásticos muestra una estructura de tipo centro-periferia.

Por tanto, en el grafo jerárquico de Ravasz se vuelve a ver la dependencia de la estructura respecto a la disposición de la matriz de adyacencia tal y como pasaba con la red telemática Internet2 y en menor medida con el conectoma de *C. elegans*. Pese a ello, lo que detecta el análisis de las matrices de adyacencia con independencia del orden de los nodos es el carácter *repetitivo* de la construcción del grafo de Ravasz (en el sentido descrito en [1]).

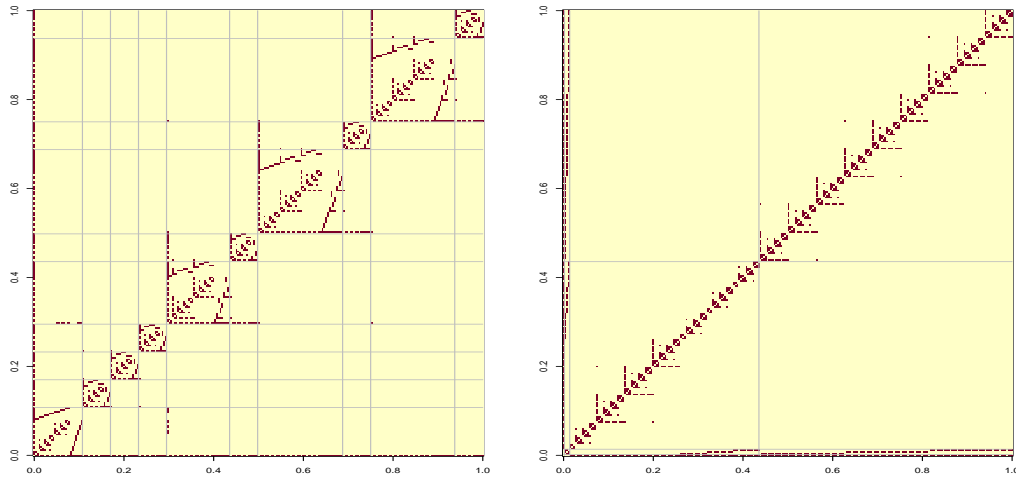


Figura 5.6: Matrices de adyacencia del grafo jerárquico. A la izquierda según el reparto modular dado por el algoritmo de Lovaina y a la derecha por el algoritmo SBM.



## Capítulo 6

# Conclusiones

En este capítulo comentamos los resultados obtenidos a lo largo de este trabajo y con ello tratamos de extraer algunas conclusiones sobre la arquitectura de las redes de neuronas.

En este trabajo tratamos de detectar la estructura comunitaria de la red asociada al conectoma de *C. elegans*. Para ello empleamos el algoritmo de Lovaina y los modelos de bloques estocásticos, obteniendo con el segundo método una distribución modular en la que se ve mejor representada la diversidad funcional de las neuronas del conectoma.

Tras esto aplicamos los dos modelos sobre las redes de control ya estudiadas, cuyas estructuras son conocidas y que por tanto nos permiten comprobar el funcionamiento de los algoritmos empleados para la detección de la estructura comunitaria. Al aplicarles el algoritmo de Lovaina obtenemos una división en módulos conexos con una alta densidad de conexiones internas y pocas conexiones con nodos externos, que es precisamente su objetivo. En cuanto a los modelos de bloques estocásticos, los resultados obtenidos para el grafo jerárquico de Ravasz parecen corroborar la idea de que el modelo del conectoma de *C. elegans* agrupa las neuronas con patrones de conectividad similares, vinculados a aspectos funcionales y espaciales, pero puramente topológicos y geométricos en el caso de la red de control. Sin embargo, el modelo de bloques estocásticos de la red telemática Internet2 carece por completo de interés, lo que concuerda con las características de la red, que no fue generada por medio de patrones repetitivos, sino que las conexiones se producen en base a la proximidad de las universidades a los conectores primarios.

En base a esto y a los resultados obtenidos en la Sección 5.3, podemos afirmar que el algoritmo de Lovaina introduce un sesgo en el procedimiento de detección de la estructura comunitaria de las redes, pues su función es encontrar una estructura asortativa con núcleos densamente conectados, por lo que la estructura comunitaria obtenida al aplicar dicho método está condicionada en cierto modo por esta perspectiva. En las dos redes de control estudiadas, el algoritmo de Lovaina encuentra comunidades que no se ajustan a la

disposición modular real de las mismas. Intenet2 no tiene una estructura asortativa, sino de tipo centro-periferia, con un conjunto de nodos densamente conectados a los que se unen nodos terminales que conforman la periferia de la red.

En cuanto al método SBM, en base a los resultados obtenidos, parece razonable pensar que los modelos determinan la estructura modular de las redes en base a la función desempeñada por sus nodos, agrupando aquellos vértices que realizan las mismas funciones en la red, pero hacen falta pruebas adicionales para asimilar los patrones de conectividad con funciones. En una red jerárquica construida mediante procesos inductivos como la de Ravasz, el algoritmo es capaz de detectar los distintos niveles jerárquicos de la red a partir de sus conexiones. Por otra parte, en la red asociada al conectoma del *C. elegans*, el algoritmo detecta ciertos patrones por los que se conectan las neuronas sensoriales y las motoras con las interneuronas y consigue vincular la estructura comunitaria con la partición funcional.

Hay autores que proponen la idea de que las redes neuronales son construcciones tipo *Lego*, donde las neuronas se unen unas a otras en bloques funcionales mediante procesos repetitivos complejos y jerarquizados. Es posible que los modelos de bloques estocásticos puedan ayudar a comprender mejor la arquitectura de las redes de neuronas. Una idea sugerente es que la combinación de los modelos de bloques estocásticos con técnicas de *deep learning* para intentar detectar patrones de conectividad en redes cerebrales hasta ahora desconocidos.

Por último, no parece posible encontrar un conjunto reducido de neuronas que conforme un núcleo funcional en las redes cerebrales. Esto se debe principalmente a la complejidad de estas redes, pues su función supone la implicación de un amplio número de neuronas y sinapsis. A pesar de que conjuntos como el *rich club* o el *hourglass weist* tienen interés para entender la topología y el flujo de información de una red de neuronas, por razones evolutivas, no es posible que el conectoma al completo dependa de ellos, pues en ese caso su falta de robustez propiciaría la extinción de la correspondiente especie.

# Bibliografía

- [1] Alcalde Cuesta F., González Sequeiros F., Lozano Rojo Á., Exploring the topological sources of robustness against invasion in biological and technological networks. *Nature Scientific Reports*, **6** (2016), 20666 EP.
- [2] Alcalde Cuesta F., González Sequeiros F., Lozano Rojo Á., A method for validating Rent's rule for technological and biological networks. *Nature Scientific Reports*, **7** (2017), 5378. doi:10.1038/s41598-017-05670-w
- [3] Bassett D. S., Greenfield D. L., Meyer-Lindenberg A., Weinberger D. R., Moore S. W. and Bullmore E. T., Efficient Physical Embedding of Topologically Complex Information Processing Networks in Brains and Computer Circuits. *PLoS Comput Biol*, **6** (2010), e1000748. doi:10.1371/journal.pcbi.1000748
- [4] Betzel R.F., Bertolero M.A. and Bassett D.S., Non-assortative community structure in resting and task-evoked functional brain networks. *BioRxiv* (2018) doi:10.1101/355016.
- [5] Clauset A., Network Analysis and Modeling, *CSCI 5352 Lecture 6* (2017).
- [6] Doyle J. C., Alderson D. L., Li L., Low S., Roughan M., Shalunov S., Tanaka R. and Willinger W., The "robust yet fragile" nature of the Internet. *Nature Scientific Reports*, **102** (2005), 14497-14502. doi:10.1073/pnas.0501426102
- [7] Good B., de Montjoye Y., and Clauset A., Performance of modularity maximization in practical contexts, *Physical review E*, **81** (2010), 046106.
- [8] Guimerà R. and Nunes Amaral L. A., Functional cartography of complex metabolic networks. *Nature*, **433** (2005), 895–900.
- [9] Hagmann P., Cammoun L., Gigandet X., Meuli R., Honey C. J., Wedeen Van J. and Sporns O., Mapping the Structural Core of Human Cerebral Cortex. *PLOS Biol*, **6** (2008), 1-15, e0060159. doi:10.1371/journal.pbio.0060159

- [10] Kim J. S. and Kaiser M., From *Caenorhabditis elegans* to the human connectome: a specific modular organization increases metabolic, functional and developmental efficiency. *The Royal Society*, **369** (2014). doi:10.1098/rstb.2013.0529
- [11] Newman M.E.J. and Girvan M., Finding and evaluating community structure in networks. *Physical review E*, **69** (2004) 026113.
- [12] Nykamp D.Q., An introduction to networks. From Math Insight. [http://mathinsight.org/network\\_introduction](http://mathinsight.org/network_introduction).
- [13] Pavlovic D. M., Vértés P. E., Bullmore E. T., Schafer W. R. and Nichols T. E. Stochastic Blockmodeling of the Modules and Core of the *Caenorhabditis elegans* Connectome. *PLOS ONE*, **9** (2014), 1-16, e0097584. doi:10.1371/journal.pone.0097584
- [14] Piñeiro Abal J., Un modelo de invasión de una red compleja: Internet2. *TFG Universidad de Santiago de Compostela*, 2017.
- [15] Ravasz E., Somera A. L., Mongru D. A., Oltvai Z. N. and Barabási A.-L., Hierarchical Organization of Modularity in Metabolic Networks. *Science*, **297** (2002), 1551–1555.
- [16] Sabrin K.M., Wei Y., van der Heuvel M.P., Dovrolis C., The hourglass organization of the *Caenorhabditis Elegans* connectome. *PLoS Comput Biol*, **16** (2020), e1007526. doi:10.1371/journal.pcbi.1007526.
- [17] Sohn Y., Choi M.-K., Ahn Y.-Y., Lee J., Jeong J.. Topological Cluster Analysis Reveals the Systemic Organization of the *Caenorhabditis elegans* Connectome. *PLoS Comput Biol*, **7** (2011), e1001139. doi:10.1371/journal.pcbi.1001139
- [18] Sporns O., Honey C.J. and Kötter R., Identification and Classification of Hubs in Brain Networks. *PLoS ONE*, **2** (2007), e1049. doi:10.1371/journal.pone.0001049.
- [19] Towlson E.K., Vértés P.E., Ahnert S.E., Schafer W.R. and Bullmore E.T., The Rich Club of the *C. Elegans* Neuronal Connectome. *The Journal of Neuroscience*, **33** (2013), 6380–6387.
- [20] Varshney L. R., Chen B. L., Paniagua E., Hall D. H. & Chklovskii D. B. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput Biol*, **7** (2011), e1001066. doi:10.1371/journal.pcbi.1001066
- [21] White J. G., Southgate E., Thomson J. N. & Brenner S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. T. R. Soc. B*, **314** (1986), 1–340.

- [22] Betzel R. F., Medaglia J. D. and Bassett D. S., Diversity of meso-scale architecture in human and non-human connectomes. *Nature Communications*, **9** (2018), 346. doi:10.1038/s41467-017-02681-z