

Victor Lázaro Vidal

"REDES MULTICAPA COMO HERRAMIENTA
DE ANÁLISIS DE LA MICROBIOTA"

2022



Universidad Autónoma de Querétaro
Facultad de Ciencias Naturales

Nombre de la tesis:

"REDES MULTICAPA COMO HERRAMIENTA DE ANÁLISIS DE
LA MICROBIOTA"

Tesis

Que como parte de los requisitos para obtener el grado de

Licenciado en Microbiología

Presenta

Víctor Lázaro Vidal

Querétaro, Qro., Mayo del 2022



Universidad Autónoma de Querétaro

Facultad de Ciencias Naturales

Licenciatura en Microbiología



Redes multicapa como herramienta de análisis de la microbiota

Tesis

Que como parte de los requisitos para obtener el grado de Licenciado en Microbiología

Presenta:

Víctor Lázaro Vidal

Dirigido por:

Dr. Roberto Carlos Álvarez Martínez

Asesores:

Dra. Karina Alethya Acevedo Whitehouse

Dr. Fausto Arellano Carbajal

Mtra. Kruskaia Caltzontzin-Fernández

Santiago de Querétaro, Qro., mayo 2022



Universidad Autónoma de Querétaro

Facultad de Ciencias Naturales

Licenciatura en Microbiología



Redes multicapa como herramienta de análisis de la microbiota

Tesis

Que como parte de los requisitos para obtener el grado de Licenciado en Microbiología

Presenta:

Víctor Lázaro Vidal

Dirigido por:

Dr. Roberto Carlos Álvarez Martínez

Asesores:

Dra. Karina Alethya Acevedo Whitehouse

Dr. Fausto Arellano Carbajal

Mtra. Kruskaia Caltzontzin-Fernández

Campus Aeropuerto

Santiago de Querétaro, Qro., México

Mayo 2022

Universidad Autónoma de Querétaro
Facultad de Ciencias Naturales
Licenciatura en Microbiología

Redes multicapa como herramienta de análisis de la microbiota

Tesis

Que como parte de los requisitos para obtener el grado de Licenciado en Microbiología

Presenta:

Víctor Lázaro Vidal

Dirigido por:

Dr. Roberto Carlos Álvarez Martínez

Dr. Roberto Carlos Álvarez Martínez

Presidente

Dra. Karina Alethya Acevedo Whitehouse

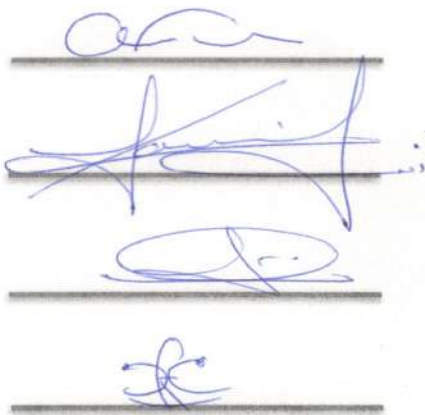
Secretaria

Dr. Fausto Arellano Carbajal

Vocal

Mtra. Kruskaia Karenia Caltzontzin Fernández

Vocal



The image shows four horizontal lines, each with a handwritten signature in blue ink above it. The signatures are stylized and cursive. The first signature is the most prominent, followed by three others that are less distinct.

Campus Aeropuerto
Santiago de Querétaro, Qro., México
Mayo 2022

Resumen

En biología existen múltiples ejemplos de sistemas complejos, es decir, sistemas conformados por una gran cantidad de componentes que interactúan entre sí, y cuyo comportamiento sólo puede entenderse como el producto de las interacciones de tales componentes. Un ejemplo de este tipo de sistemas es la microbiota, donde distintos microorganismos interactúan entre sí, con los tejidos del hospedero y con su entorno. Las redes son una herramienta matemática sumamente útil en el análisis de sistemas complejos, sin embargo, el enfoque tradicional enfrenta ciertas limitaciones, pues los sistemas biológicos son sistemas dinámicos que cambian en el tiempo y en función de las condiciones. Ante tales retos, las redes multicapa se presentan como un nuevo paradigma en el análisis de sistemas complejos. Empleando redes multiplex, analizamos bases de datos obtenidas a partir de muestras de la microbiota de dos organismos estrechamente relacionados, la solanácea *Datura innoxia* y el escarabajo *Lema daturaphila*. Las redes de coabundancia fueron generadas mediante los algoritmos de inferencia SparCC y ARACNe. Se hallaron comunidades bacterianas involucradas en la fijación de nitrógeno en la capa de la solanácea, mismas que podrían tener una relación mutualista con los tejidos de la planta. Tanto en la solanácea como en el escarabajo se encontraron comunidades bacterianas asociadas al tejido intestinal del insecto, lo que en la solanácea podría derivarse de la materia fecal depositada por las larvas mientras se alimentan de la planta. Bajo el enfoque de las redes temporales, se analizaron datos de dos sujetos que experimentaron cambios en su microbiota derivados de procesos infecciosos. Las redes ARACNe de ambos sujetos mostraron alteraciones en la arquitectura de las redes durante los periodos de disbiosis asociados a procesos infecciosos, lo que podría indicar alteraciones en la dinámica de la microbiota. También se identificaron proporciones alteradas de los *phyla* durante los estadios de disbiosis. Además, gran parte de los nodos encontrados en una misma comunidad eran filogenéticamente cercanos. El uso de redes multicapa nos permitió encontrar patrones en las comunidades bacterianas que podrían estar asociados a procesos puntuales, como la liberación de AGCCs y el metabolismo de determinados nutrientes.

Summary

In biology there are multiple examples of complex systems. It means systems made up of many components that interact with each other, and whose behavior can only be understood as the sum of the interactions of its components. An example of this type of system is the microbiota, where different microorganisms interact with each other, with the host's tissues, and with their environment. Networks are a handy mathematical tool for analyzing complex systems; however, the traditional approach faces certain limitations because biological systems are dynamical systems that change over time and conditions. In order to deal with such challenges, multilayer networks appear as a new paradigm in analyzing complex systems. Through the use of multiplex networks, I analyzed databases obtained from samples of the microbiota of two closely related organisms: the nightshade *Datura innoxia* and the beetle *Lema daturaphila*. I use SparCC and ARACNe algorithms for inferring co-abundance networks. Bacterial

communities involved in nitrogen fixation in the solanaceous layer abound, which could have a mutualistic relationship with plant tissues. Bacterial communities associated with the insect's intestinal tissue appeared on both layers, something that could derive from the larvae' dregs in solanaceous while feeding on the plant. Under the approach of the temporal network, I analyzed data from two subjects who experienced changes in their microbiota derived from infectious processes. The ARACNe networks of both subjects showed alterations in the architecture of the networks during periods of dysbiosis associated with infectious processes, which could indicate alterations in the dynamics of the microbiota. Altered *phyla* proportions came into sight during dysbiosis stages. Furthermore, most of the nodes found in the same community were phylogenetically close. The use of multilayer networks allowed us to find patterns in bacterial communities that could be associated with specific processes, such as the release of SCFAs and the metabolism of certain nutrients.

Dedicatoria

A mi mamá, quien siempre me ha acompañado, socorrido, apoyado e instruido a lo largo de mi vida, y a quién dedico con especial aprecio este trabajo y todo cuanto logre en la vida.

A mi hermana, quién ha sido mi compañera y amiga, y la persona con la que he crecido y convivido desde niño.

A mi papá, que, aunque ya no esté para verlo, espero se sienta orgulloso de lo que he logrado hasta ahora.

A mi abuelita, quien siempre me ha querido y cuidado, y con quien he compartido algunos de mis mejores recuerdos.

Agradecimientos

Ahora que he concluido esta etapa de mi vida, espero poder continuar mi formación y contribuir a la sociedad, al mundo en el que vivo y sobre todo a mi familia y amigos, a todas las personas que aprecio y a las cuales agradezco su contribución no solo en mi formación académica, sino también en mi formación como persona. Y a pesar de que siga siendo en extremo distraído y abstraído de la realidad, espero poder dejar una pequeña huella, especialmente en la memoria de aquellas personas a las que quiero.

En primero lugar me gustaría agradecer a mi director de tesis, el Dr. Roberto, quien, a pesar de todos mis tropiezos y mi constante tendencia a distraerme, me ha apoyado y acompañado en todo este camino, motivado y ayudado cuando lo he necesitado. También me gustaría agradecer a todos los profesores que, en mayor o menor medida, han contribuido a mi formación, no solo en la universidad sino también desde mucho antes. Del mismo modo agradezco a mi familia, incluyendo tíos y primos, y especialmente a mi madre, hermana, padre y abuela, quienes han estado siempre allí para mí y espero estar del mismo modo para ellos. A mis amigos y compañeros que, de una u otra forma, han contribuido a quien soy ahora: A Erick, Jorgi, Brandon, Isabel, Dany, Giss y Ale.

Tabla de contenido

Resumen

Summary

Agradecimientos

Tabla de contenidos

Índice de cuadros y figuras

Figuras

Tablas

1. Introducción

2. Antecedentes

2.1 Definición y generalidades de una red

2.2 Propiedades topológicas de las redes

2.3 Definición y características de las redes multicapa

2.4 Aplicaciones de las redes multicapa

2.5 Inferencias de redes

2.6 La microbiota intestinal y su impacto en la salud humana

2.7 Interacción planta-microbiota-insecto

2.8 Uso de redes en la exploración de la microbiota

3. Objetivo general

3.1 Objetivos particulares

4. Materiales y métodos

5. Resultados

5.1 Resultados de la red escarabajo-solanácea

5.2 Resultados de las redes para los sujetos A y B

6. Discusión

6.1 Discusión de la red escarabajo-solanácea

6.2 Discusión de las redes temporales de los sujetos A y B

6.3 Conclusiones generales

9. Referencias

Índice de cuadros y figuras

Figuras

Figura 1: Red aleatoria y red de mundo pequeño [1].

Figura 2: La aleatoriedad de las redes [2].

Figura 3: Aleatoriedad de la red vs. longitud de la red [2].

Figura 4: Tipos de centralidades [3].

Figura 5: Tipos de redes multicapa [4].

Figura 6: Red multiplex de familias florentinas del Renacimiento [5].

Figura 7: Matriz de supra-adyacencia [6].

Figura 8: Coeficiente de clusterización en redes multicapa [7].

Figura 9: Análisis de las propiedades de una red multiplex del metro de Zaragoza [8].

Figura 10: Red temporal de un modelo SIS [9].

Figura 11: Diagrama de interacciones del genoma, epigenoma, transcriptoma, proteoma y metaboloma [10].

Figura 12: Diagrama de interacciones en un ecosistema [11].

Figura 13: Funcionamiento de los algoritmos de inferencia de redes [12].

Figura 14: Inferencia mediante el algoritmo WGCNA [13].

Figura 15: Ejemplo de conexiones directas e indirectas.

Figura 16: Curvas PR para algoritmos CRL, ARACNe y MRNet [14].

Figura 17: Comparación entre la inferencia de Pearson y SparCC [15].

Figura 18: Inferencia de redes por SparCC [16].

Figura 19: Discrepancia de la inferencia entre varios algoritmos [17].

Figura 20: Interacción entre la microbiota y el huésped bajo distintos regímenes alimenticios [18].

Figura 21: Efectos de la disbiosis en el tejido hepático [19].

Figura 22: Redes de coabundancia de la microbiota de individuos con síndrome de intestino irritable [20].

Figura 23: Red multicapa de la microbiota del rumen bovino [21].

Figura 24: Clusterización mediante el método Louvain de las capas de la red escarabajo-solanácea.

Figura 25: Red multiplex SparCC de escarabajo-solanácea.

Figura 26: Distribución de los *phylum* en las muestras de la red escarabajo-solanácea.

Figura 27: Centralidad por *degree* de la red escarabajo-solanácea.

Figura 28: Cambio de las abundancias de los nodos entre capas de la red escarabajo-solanácea.

Figura 29: Distribución de las abundancias de los *phyla* en el sujeto A a lo largo de las tres capas.

Figura 30: Capas de las redes temporales ARACNe de los sujetos A y B.

Figura 31: Serie temporal de los diez OTUs más abundantes en las muestras del sujeto A.

Cuadros

Cuadro 1: Paquetes de R utilizados.

Cuadro 2: Funciones del paquete ml_BioNets.

1. Introducción

Existe un debate en cuanto al estudio de los sistemas físicos y su comportamiento, uno que involucra a dos posturas aparentemente contrapuestas. Por un lado, se plantea que las leyes fundamentales son suficientes para describir el comportamiento de cualquier sistema. La hipótesis reduccionista estipula que un sistema puede ser entendido mediante el estudio de sus partes fundamentales, una idea que parte de la concepción de que todo dentro de nuestro universo está regido por las mismas leyes físicas [22]. Sin embargo, este pensamiento resulta insuficiente cuando, a partir de la interacción de los distintos componentes de un sistema, surgen fenómenos que no se pueden predecir basándose únicamente en las leyes fundamentales que rigen sus componentes, a lo cual llamamos propiedades emergentes [23].

Al hablar de sistemas complejos y basándose únicamente en las leyes fundamentales de la física, química y biología, resulta problemático predecir la diversidad de comportamientos que observamos en los organismos vivos [24]. Definimos una propiedad emergente como un fenómeno producto de la interacción de un conjunto de entidades simples que operan en un entorno, dando lugar a comportamientos colectivos más complejos [25]. En biología, los ejemplos de sistemas complejos son abundantes, como la forma en que distintas biomoléculas interactúan dentro de una célula para conducir señales, las conexiones entre las neuronas y la regulación de la respuesta inmune en el organismo, la propagación de enfermedades infecciosas, la comunicación entre insectos sociales y las relaciones entre las distintas especies que componen un ecosistema, etc. [26].

Una red consiste en un conjunto de nodos o elementos que a su vez están conectados por aristas [27]. Las redes de sistemas biológicos a menudo presentan particularidades en su arquitectura, como, por ejemplo, el hecho de que las redes de señalización de una célula sean robustas, donde unos pocos elementos mantienen la mayoría de las conexiones, hace que la red sea poco susceptible a perder funcionalidad ante ataques aleatorios a su estructura [28]. Es particularmente importante, por ejemplo, si hablamos de sistemas de regulación génica dirigidos a mantener el funcionamiento de una célula, pues será más probable que un ataque no-dirigido afecte a un componente no esencial de la red, a que interfiera con un elemento que sí lo sea [28].

Las redes multicapa pueden ser utilizadas para analizar la forma en la cual distintos sistemas interactúan o la forma en la que estos evolucionan en el tiempo [4]. Uno de estos enfoques son las redes temporales, donde cada capa corresponde a un estado de la red en una fracción determinada de tiempo, por lo que permite visualizar cómo cambia la estructura de dicha red con respecto al tiempo [4]. Este tipo de enfoque es particularmente útil, por ejemplo, para predecir el comportamiento de una epidemia en una población a pasos discretos, pues las interacciones sociales cambian en el tiempo, modificando la forma en que se propaga el agente infeccioso [29]. Además, existe el caso de sistemas que pueden interactuar a distintos niveles, como los mecanismos de regulación de una célula, donde proteínas pueden interactuar directamente con el DNA o RNAs que pueden, a su vez, bloquear el procesamiento de transcritos. Esto significa que cada una de las capas de la multired puede interactuar con cualquiera de las otras capas y no exclusivamente con las capas contiguas [30].

Por su puesto, la microbiota no es la excepción, e investigaciones recientes han revelado que esta es parte de una extensa red que posee un efecto no sólo sobre las mismas poblaciones bacterianas, sino también a nivel de células y tejidos del hospedero, pudiendo tener un impacto importante sobre su fisiología e inclusive sobre la conducta [31]. El estudio de la microbiota mediante el enfoque de las redes multicapa ha ayudado a comprender el papel e importancia de esta en la salud, así encontrar patrones asociados a enfermedades metabólicas y otros padecimientos [31].

2. Antecedentes

2.1 Definición y generalidades de una red

Una red se define como un conjunto de elementos o nodos, que denotaré como V , que se conectan entre sí mediante aristas representadas por el conjunto E [4]. $G = (V, E)$, donde $N = |V|$ corresponde al conjunto de nodos que componen la red, y $L = |E|$ al conjunto de aristas que conectan los nodos o vértices [4].

$$V = \{i | i \in \{1, 2, \dots, N\}\}$$

La información de las conexiones entre los nodos se contiene dentro de una matriz de adyacencia $N \times N$, donde se coloca un valor de 1 por cada interacción entre un par de nodos, y de 0 si no existe interacción entre nodos [4]. En matrices no-dirigidas las matrices son especulares, mientras que en matrices dirigidas esta propiedad se pierde [32].

$$a_{ij} = \begin{cases} 1 & \text{si } i \leftrightarrow j \\ 0 & \text{si no.} \end{cases}$$

	α_1	α_2	α_3	...	α_{n-1}	α_n
β_1	0	1	0	...	1	1
β_2	1	0	1	...	0	1
β_3	0	1	0	...	1	0
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
β_{n-1}	1	0	1	...	0	1
β_n	1	1	0	...	1	0

2.2 Propiedades topológicas de las redes

Cada nodo de la red posee un determinado número de conexiones con otros nodos, una característica que se conoce como *degree*, y que se define como la sumatoria de todas las interacciones de un nodo con otros nodos de la red [33].

$$k_i = \sum_{j=1}^N a_{ij}$$

Las interacciones entre estos nodos pueden estar o no dirigidas, así como estar o no ponderadas, como veremos a continuación. Se considera que las conexiones de una red son ponderadas, es decir, que poseen un valor numérico asociado a la fuerza de la interacción [33]. En sistemas biológicos, por ejemplo, esta ponderación podría representarse como la correlación entre grado de expresión de dos genes o la abundancia relativa de dos poblaciones bacterianas [4].

$$k_i = \sum_{j=1}^N \theta(a_{ij})$$

Las conexiones de las redes pueden ser dirigidas, es decir, la conexión del nodo i con j no establece de forma implícita la conexión del nodo j al i [4]. Algunos ejemplos de redes no-dirigidas son las redes de regulación transcripcional o las redes de regulación por microRNA-transcritos, donde si un elemento x está regulando positiva o negativamente a un elemento y , no implica que y también regule a x .

Las redes no-dirigidas asumen que la interacción es bidireccional, algo que no ocurre necesariamente en las redes dirigidas, por lo que en este segundo caso la matriz de adyacencia es asimétrica [4].

$$k_{i,in} = \sum_{j=1}^N \theta(a_{ji})$$

$$k_{i,out} = \sum_{j=1}^N \theta(a_{ij})$$

El coeficiente de clusterización o de transitividad cuantifica la densidad de triángulos que se forman entre los vecinos de un nodo i , lo cual se puede interpretar como la probabilidad de que un nodo esté interconectado con otros [34]. Para una red no dirigida, el coeficiente de clusterización se calcula de la siguiente forma.

$$C_i = \begin{cases} \frac{2 \cdot (\text{numero de triangulos pasando por el nodo } i)}{k_i(k_i-1)} & \text{para } k_i > 1 \\ 0 & \text{para } k_i = 1 \text{ o } 0 \end{cases}$$

Un coeficiente de transitividad igual a 0 significa que los nodos que interactúan con el nodo i , no están conectados entre sí, por lo que este valor aumenta entre mayor sea la conectividad de los nodos vecinos [1]. Un coeficiente de transitividad cercano a 1 implica que los nodos están estrechamente conectados en la región analizada, y, por lo tanto, que conforman un clúster o conjunto dentro de la red [1]. En sistemas biológicos, como en una red de distintos genes, regiones con un coeficiente de clusterización alto podría indicar que los genes analizados estarían posiblemente involucrados en un proceso biológico común [30].

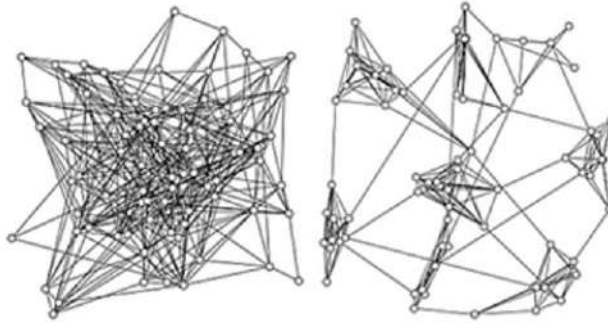


Figura 1: Las redes generadas aleatoriamente (izquierda) tienden a tener un coeficiente de clusterización significativamente menor al de las llamadas redes de mundo pequeño (derecha). Las redes de mundo pequeño se caracterizan por tener conjuntos de nodos estrechamente interconectados, a pesar de que la mayoría de sus elementos no sean vecinos entre sí. Imagen extraída de Schaeffer, 2007 [1].

El coeficiente de clusterización es una propiedad local, es decir, para cada nodo se define un valor de dicho coeficiente. Para determinar un valor global del coeficiente de clusterización de la red en su conjunto es simplemente el promedio aritmético de todos los coeficientes de clusterización.

$$C_{net} = \frac{1}{N} \sum_{i=1}^N C_i$$

Al tipo de redes donde la mayoría de los nodos no son vecinos entre sí se les conoce como redes de mundo pequeño. Los primeros en idear este coeficiente fueron Watts y Strogatz en 1988, y es gracias a esta propiedad que es posible diferenciar una red aleatoria de una red de mundo pequeño [2].

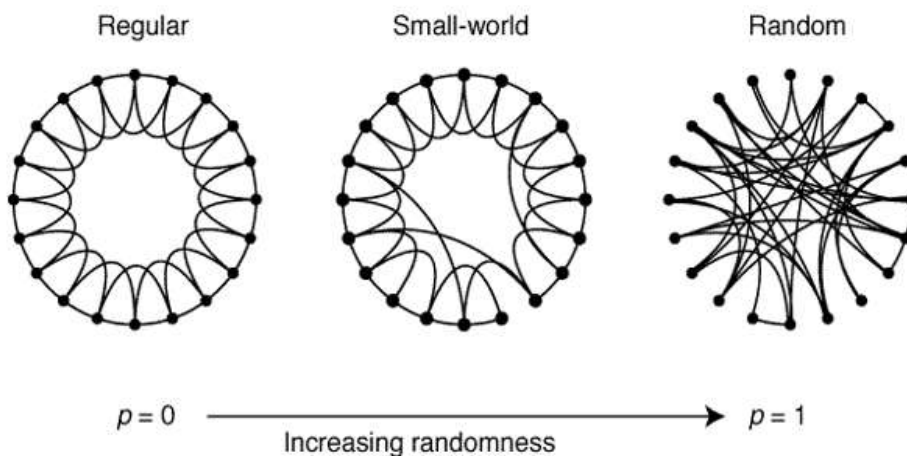


Figura 2: La aleatoriedad de una red (p) es un factor importante en su arquitectura. Cuando este valor es igual a cero, los nodos poseerán exactamente el mismo número de conexiones y, por lo tanto, la distribución de su degree será uniforme. Por el contrario, cuando los valores de aleatoriedad son cercanos a uno, dicha distribución se asemeja a una curva gaussiana, donde existirán nodos que posean muchas conexiones en un extremo de la distribución, otros con un número bajo de conexiones y la mayoría con un número de conexiones cercano a la media de la red. Imagen extraída de Watts y Strogatz, 1998 [2].

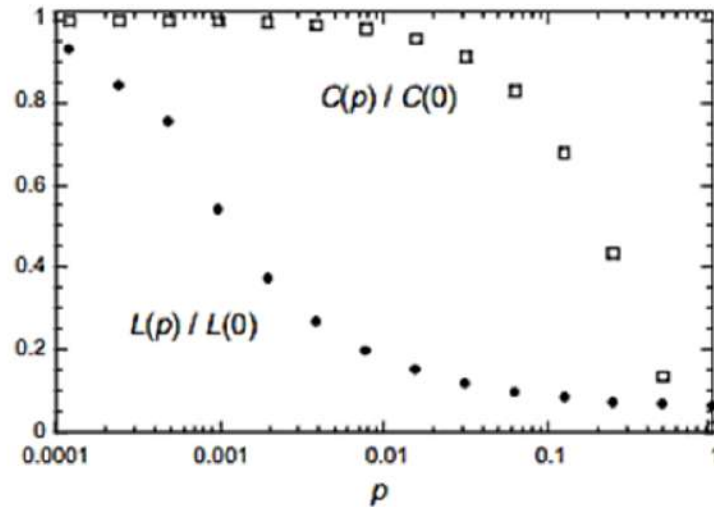


Figura 3: Conforme aumenta la aleatoriedad de una red, el coeficiente de clusterización ($C(\rho)$) adquiere valores cada vez menores; al tiempo que la longitud de la red ($L(\rho)$), es decir, la distancia (cuantificada en número de aristas) más corta entre todos los posibles pares de nodos de una red, también disminuye. Imagen extraída de Watts y Strogatz, 1998 [2].

Esta característica le brinda a la red robustez, es decir, la capacidad de conservar su arquitectura ante ataques aleatorios [35]. En sistemas biológicos, como pueden ser cascadas de señalización, de suceder mutaciones aleatorias que puedan alterar el funcionamiento normal de alguna proteína involucrada en determinada ruta de señalización, es estadísticamente más probable que dicha mutación ocurra en un elemento de la red con bajo número de conexiones a que suceda en algún elemento central de la red, lo cual permite que la célula pueda continuar sus funciones biológicas con relativa normalidad [36]. En este tipo de redes unos pocos nodos poseen la mayor parte de las conexiones, mientras que la mayoría de los nodos poseen un bajo número de estas [35].

Existen diversas formas de conocer la centralidad o importancia de un nodo dentro de una red. La centralidad por *degree* o grado se fundamenta en el número de conexiones de los nodos, mientras que la centralidad por intermediación o *betweenness* otorga mayor peso a nodos que actúan como puentes entre clústers separados. La centralidad por *closeness* o cercanía cuantifica el promedio de las distancias más cortas de cada nodo con todos los demás. La centralidad por *eigenvectores*, también conocida como *eigencentality*, es una medida autorreferencial de centralidad que prima la calidad de las conexiones, donde un nodo tendrá mayor relevancia si a su vez está conectado a un nodo con alta centralidad. El coeficiente de partición, por su parte, considera la distribución de las conexiones entre clústers separados. Finalmente, el *PageRank* es una medida de centralidad similar al *degree* pero empleada en redes dirigidas, y a menudo usada para determinar la importancia de páginas web [3].

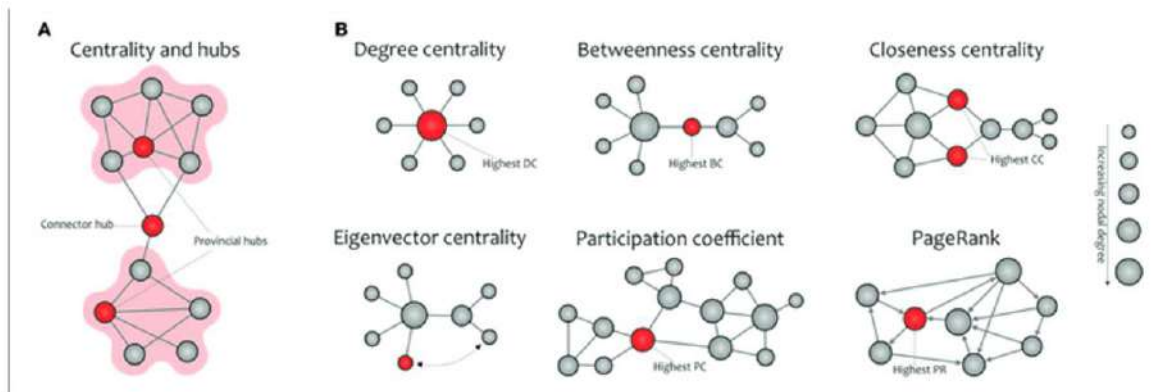


Figura 4: A) Los *hubs* son nodos con un alto valor de centralidad. B) Medidas de centralidad. Imagen extraída de Farahani et al., 2019 [3].

Sin embargo, los sistemas biológicos no suelen ser entes aislados, sino que se ven afectados por su interacción con aquello que los rodea. La idea de que los componentes de una red son capaces de interactuar o relacionarse de alguna forma con los elementos de otras redes, nos adentra en el concepto de las redes múltiples (*multinetworks*), también conocidos como multigrafos (*multigraphs*) o redes multicapa (*multilayer networks*).

2.3 Definición y características de las redes multicapa

Una red multicapa es una red compuesta por M redes individuales cuyos elementos interactúan entre sí, constituyendo cada una de estas redes individuales una capa [37].

$$\mathcal{M} = (Y, \vec{G}, \mathcal{G})$$

Donde Y corresponde al conjunto de capas:

$$Y = \{\alpha | \alpha \in \{1, 2, \dots, M\}\}$$

\vec{G} corresponde a la lista ordenada de redes que conforman cada una de las capas, donde cada elemento $G_\alpha = (V_\alpha, E_\alpha)$ es una red monocapa con sus propios nodos y aristas dentro de la red multicapa [37].

$$\vec{G} = (G_1, G_2, \dots, G_\alpha, \dots, G_M)$$

A continuación, la matriz $M \times M$, en la cual están contenidas las interacciones entre los nodos de cada par de redes que conforman la red multicapa. Debido a que en la matriz están representadas las interacciones entre cada par de redes, podemos entender el total como un conjunto de redes bipartitas [37].

$$\mathcal{G} = (V_\alpha, V_\beta, E_{\alpha,\beta})$$

Las redes multicapa se clasifican en tres categorías principales: las redes multiplex, redes temporales y redes formadas por redes [4]. Las redes multiplex son el ejemplo más sencillo de redes multicapa; estas se emplean generalmente cuando el mismo conjunto de nodos se encuentra contenido en las distintas capas, pero la interacción entre estos es distinta en cada una [4]. Las redes multiplex son empleadas para mostrar cómo un mismo conjunto de elementos pueden interactuar entre sí de distintas formas, por lo general, asociando las aristas con un color distinto para cada tipo de interacción [38]. Por ejemplo, podríamos ejemplificar esto como un conjunto de genes que se relacionan entre sí en distintas células [30].

Por su parte, los sistemas dinámicos son aquellos donde las interacciones cambian constantemente en función del tiempo y, por lo tanto, la forma en que sus componentes están relacionados depende en gran medida del momento y las condiciones en las cuales se realiza la medición [39]. Los sistemas biológicos son por lo general sistemas dinámicos, pues estos están constantemente sometidos a cambios debidos a estímulos con el ambiente que los rodea, así como a la naturaleza cambiante de los procesos que ocurren dentro de las células. Por ejemplo, un conjunto de genes no permanecerá siempre expresado de la misma manera, sino que el grado de expresión de estos dependerá de un gran número de estímulos. Este tipo de sistemas se pueden representar mediante redes temporales, donde cada una de las capas representa una captura de las interacciones entre un conjunto de nodos en un momento determinado [39]. Las redes temporales son secuenciales, es decir, que las capas están ordenadas de modo que se suceden una detrás de otra, por lo cual una capa t_i solo puede interactuar con la capa siguiente t_{i+1} [40]. Este tipo de redes son similares a las redes multiplex en el sentido de que se trata del mismo conjunto de nodos entre las capas [40].

Finalmente podemos destacar las redes formadas por redes. Estas poseen importantes diferencias con respecto a los tipos de redes ya mencionados, empezando con que los nodos varían entre capa y capa; además de que los nodos de una capa no interactúan necesariamente solo con las capas adyacentes, como sí sucede con las redes multiplex o las redes temporales, sino que pueden interactuar con cualquier número de nodos en cualquiera de las otras capas [41].

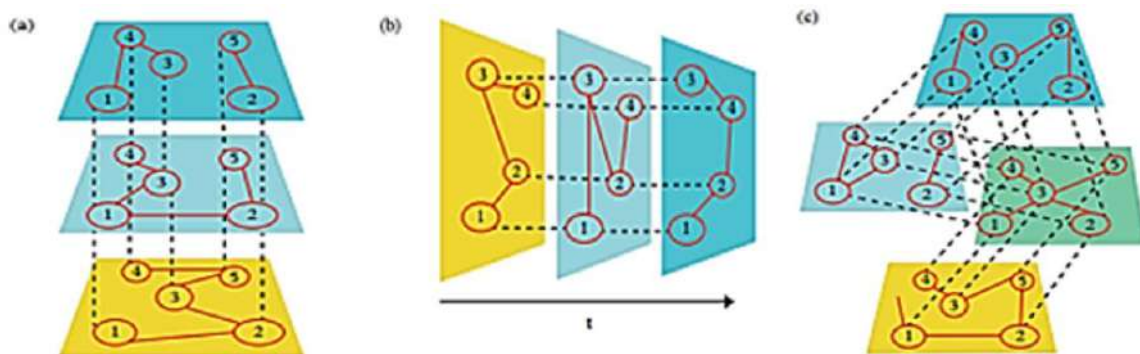


Figura 5: Existen distintos tipos de redes multicapa, de las cuales podemos resaltar tres categorías: a) Redes multiplex. b) Redes temporales. c) Redes de redes. Imagen extraída de Bianconi, 2018 [4].

Debido a que las interacciones entre capas de una red multiplex están limitadas a las capas vecinas con respecto a la capa de referencia, podemos representar a una red múltiplex como una red bidimensional con conexiones de distintos tipos. En este tipo de redes los nodos de una capa solo se conectan con sus homólogos en la siguiente, además de que el conjunto de estos se mantiene en las distintas capas, variando solo las interacciones entre los nodos dentro de las mismas capas [38].

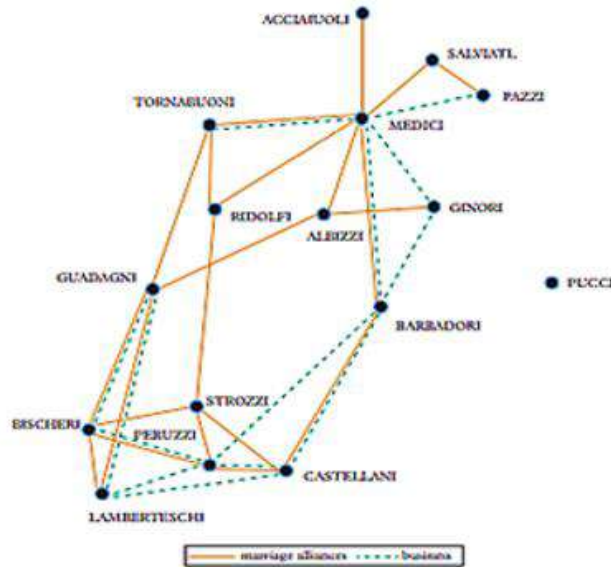


Figura 6: Las redes multiplex muestran cómo los nodos de una red pueden interactuar entre sí de distintas formas. En este ejemplo se muestran los vínculos matrimoniales y de negocios entre distintas familias nobles florentinas del siglo XV. Utilizando esta representación, podemos vislumbrar con facilidad el papel central que tenía la familia Medici tanto en el ámbito comercial como político entre la nobleza florentina. Imagen extraída de Biaconti, 2018 [5].

$$\mathcal{M} = (Y, \vec{G})$$

Donde

$$\vec{G} = (G_1, G_2, \dots, G_\alpha, \dots, G_M)$$

Cada una de las capas de la red multiplex poseen la misma cantidad de nodos

$$V = \{i | i \in \{1, 2, \dots, N\}\}$$

Sin embargo, a pesar de que las redes multiplex presentan el mismo grupo de nodos en las distintas capas, a menudo es necesario distinguir la capa en la cual se encuentra cada uno de los nodos [4].

$$V = \{i_\alpha | i \in \{1, 2, \dots, N\}\}$$

Tanto para redes múltiplex como para redes temporales, la conexión de los nodos entre pares de capas está dada por la delta de Kronecker. Ésta es una función cuya variable

puede adquirir un valor de 1 en aquellos casos donde un elemento sea igual en ambos conjuntos de nodos, o en este caso en dos capas, y de 0 cuando son diferentes. Aquellos nodos que cumplen la condición de ser equivalentes entre una capa y otra se denominan nodos réplica [42].

$$V = \{i_\alpha | i \in \{1, 2, \dots, N\}\}$$

Como hemos visto previamente, la información de una red simple está contenida dentro de una matriz de adyacencia, sin embargo, en redes multicapa este modelo resulta insuficiente para contener la información de estas estructuras, razón por la cual empleamos las llamadas matrices de supra-adyacencia. Este tipo de matrices están compuestas a partir de múltiples matrices de adyacencia individuales, y contiene tanto la información referente a las interacciones dentro de las capas como las interacciones entre capas [6].

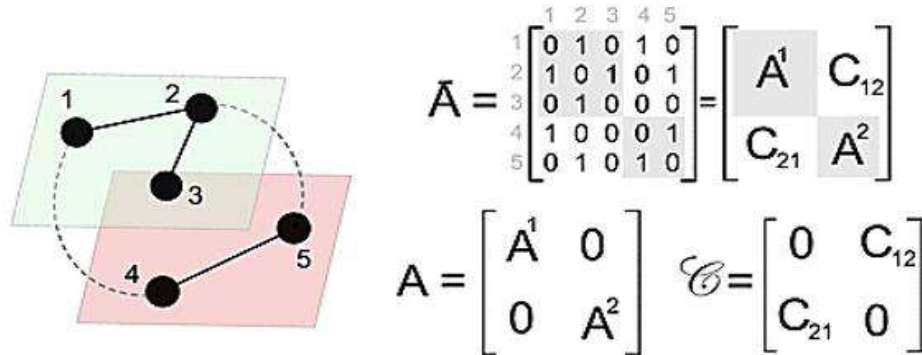


Figura 7: A la derecha se muestra la representación tridimensional de una red bipartita. La matriz de supra-adyacencia \bar{A} está compuesta a partir de la fusión de dos matrices de adyacencia simples. La primera es la matriz A , que codifica las interacciones de los nodos dentro de las capas, mientras que la matriz C contiene las interacciones entre las capas. Imagen extraída de Cozzo et al., 2016 [6].

Como se puede inferir, las propiedades de las redes múltiples difieren de las redes simples. Como ya hemos visto, el *degree* (k_i) de un nodo de una red simple es un escalar, un número entero positivo que representa el número de conexiones de un nodo. En redes múltiples, este valor está representado en un vector M -dimensional que contiene los valores del *degree* de un nodo “ i ” a lo largo de las M capas de la red [43].

$$k_i = (k_i^{[1]}, k_i^{[2]}, \dots, k_i^{[M]})$$

Al igual que como sucede con redes simples, la distribución del *degree* es una función de probabilidad $P^{[a]}(k)$. En redes múltiples, esta función de probabilidad se define de forma independiente para cada capa [43].

$$P^{[a]}(k) = \frac{N^{[a]}(k)}{N_a}$$

El coeficiente de clusterización en una red multiplex, al igual que sucede con las redes simples, permite conocer si existen conjuntos donde varios nodos están

interconectados entre sí, sólo que, en este caso, esto también incluye las conexiones que se generan entre las capas [44]. En redes sociales, por ejemplo, esta propiedad permitirá conocer la probabilidad de que dos individuos realicen varias actividades juntos, o que dos personas en dos clases distintas tengan algún amigo en común [4]. En biología, por ejemplo, el coeficiente de clusterización podría emplearse para saber si dos genes que se sobreexpresan en distintos pacientes analizados, podrían estar involucrados en el mismo proceso [29]. El problema de la clusterización en redes múltiples es que los ciclos de longitud, los mismos que generan triángulos de nodos interconectados en redes simples, se completan en capas distintas, lo que complica su identificación [44].

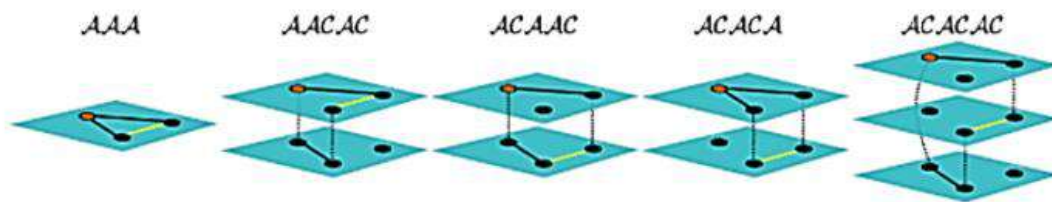


Figura 8: En el extremo izquierdo de la imagen se muestra una red simple donde los dos vecinos del nodo i (marcado en rojo), están conectados entre sí. En los siguientes tres diagramas se muestran ejemplos de redes bipartitas donde, con respecto a la capa donde se encuentra el nodo i , un grupo de nodos pueden no ser vecinos en dicha capa, pero si en una capa distinta, permitiendo la formación de ciclos entre ambas capas. En el extremo derecho de la imagen podemos ver un ejemplo en el cual se ejemplifica cómo estos ciclos de nodos interconectados no necesariamente se limitan a dos capas, sino que pueden extenderse a más capas. Imagen extraída de Cozzo et al., 2013 [7].

De forma similar a como hemos visto en redes simples, podemos obtener este coeficiente dividiendo el número de triángulos que se forman entre los nodos de distintas capas sobre el producto del número de vecinos (*degree*) de la primera y última capa ($N_{\alpha,\gamma}$) [7].

$$C_i^{[\alpha,\beta,\gamma]} = \frac{\sum_{r,s} a_{ir}^{[\alpha]} a_{rs}^{[\beta]} a_{si}^{[\gamma]}}{N_{\alpha,\gamma}}$$

Donde:

$$N_{\alpha,\gamma} = \begin{cases} k_i^{[\alpha]} k_i^{[\gamma]} & \text{para } \alpha \neq \gamma \\ k_i^{[\alpha]} (k_i^{[\alpha]} - 1) & \text{para } \alpha = \gamma \end{cases}$$

Así como sucede en lo referente a la transitividad, la cuantificación de otras propiedades de la red, tales como las distancias entre los nodos, difieren cuando tratamos con redes multicapa. Inferir la distancia más corta entre cualquier par de nodos implica ahora considerar las distancias entre capas (ψ_{ij}), además de las distancias más cortas dentro de la misma capa (σ_{ij}). La fracción resultante nos permite

conocer la accesibilidad entre los nodos, una propiedad a la cual denominamos interdependencia (λ_i), cuyo valor fluctúa entre 1 y 0 [45].

$$\lambda_i = \sum_{i \neq j} \frac{\psi_{ij}}{\sigma_{ij}}$$

La interdependencia media de la red multicapa puede expresarse como:

$$\lambda = \frac{1}{N} \sum_{i=1}^N \lambda_i$$

Cuando λ posee un valor equivalente o muy cercano a 0, significa que existe una muy baja interdependencia entre las capas, ya que todas o la gran mayoría de las rutas más cortas entre cualquier par de nodos se encuentran dentro de la misma capa. Lo contrario sucede cuando λ adquiere valores cercanos o equivalentes a 1, la interdependencia de los nodos entre las capas es muy alta [45].

El análisis de las propiedades de la red puede arrojar indicios sobre el comportamiento del sistema en su conjunto, así como información que ayude a explicar ciertos fenómenos que ocurren cuando estudiamos ejemplos reales de sistemas complejos.

2.4 Aplicaciones de las redes multicapa

Las redes multicapa representan un enfoque relativamente nuevo en teoría de redes y en el análisis de sistemas complejos. Los sistemas complejos son aquellos que están compuestos por un número muy grande de distintos elementos que interactúan, de forma no lineal, entre sí [46]. A medida que el estudio de este tipo de sistemas avanzaba, fue evidente pronto que no se trataba de sistemas aislados, y que, por el contrario, se requería de una nueva perspectiva que ampliara nuestra capacidad para analizarlos, dando lugar al enfoque de las redes multicapa [4]. Las redes multicapa son ubicuas, podemos encontrar ejemplos en casi todo aquello que nos rodea; desde la evolución de las conexiones neuronales con respecto al tiempo, las interacciones entre el proteoma, el genoma y el transcriptoma, hasta el internet, los medios de transporte urbano en una ciudad o las interacciones entre las distintas partículas elementales [4].

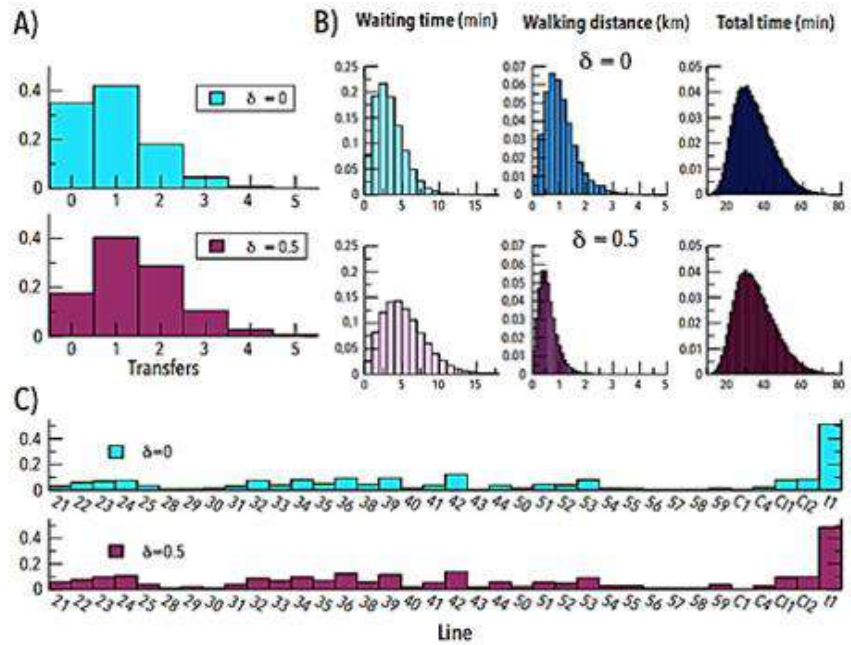


Figura 9: Mediante el análisis de las propiedades de las redes múltiples a gran escala es posible obtener información sobre el comportamiento de un sistema. En el ejemplo se realizó la simulación de un modelo basado en el sistema de transporte de Zaragoza, España. a) Distribución del número de veces que los individuos pudieron llegar a su destino en el intervalo de tiempo en que se realizó la simulación. b) Distribución del tiempo de espera, la distancia recorrida y la duración del recorrido empleando dos velocidades de movimiento distintas (δ). c) Fracción de pasajeros que usaron las distintas líneas de transporte. Imagen extraída de Aleta et al., 2017 [8].

Aunque las redes multicapa se estudiaron originalmente en el contexto de las ciencias sociales con el fin de representar la forma en que distintas personas establecen lazos de formas diferentes, por ejemplo, a través de las relaciones laborales, de amistad o mediante distintas redes sociales [47]; el uso de este enfoque se ha extendido a muchos otros campos, y las ciencias biológicas no han sido la excepción. Desde la neurociencia y la biología molecular hasta la ecología y la epidemiología, las redes multicapa han tenido multitud de aplicaciones dentro del campo de la biología y la medicina.

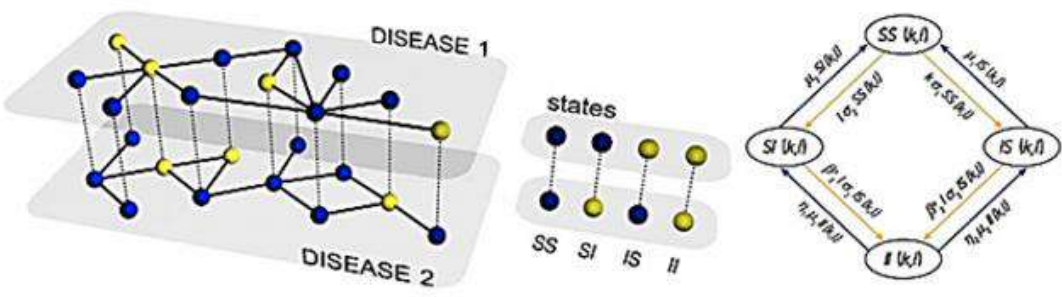


Figura 10: Una de las aplicaciones más recurrentes en biología de las redes multicapa es en el campo de la epidemiología, en donde se aplican redes temporales para predecir el comportamiento de un brote infeccioso

en una población, como se muestra en esta representación de un modelo SIS. Imagen extraída de Sanz et al., 2014 [9].

Para el caso de las redes de redes podemos aludir al ejemplo de genes que interactúan entre ellos y que a su vez producen proteínas cuya interacción genera un efecto en las células, a la vez que regulan el sistema a distintos niveles: genoma, epigenoma, transcriptoma, proteoma y metaboloma. A este conjunto de biomoléculas en interacción que constituyen una célula es a lo que llamamos interactoma [30]. Una proteína producida por determinado gen puede regular positiva o negativamente a otro gen al realizar una modificación epigenética sobre este, así como un siRNA es capaz de regular la traducción a proteína de determinado transcrito. Como podemos ver con este ejemplo, la regulación de la red puede ocurrir a distintos niveles y mediante la interacción de elementos de alguna capa con cualquiera de las otras capas [10].

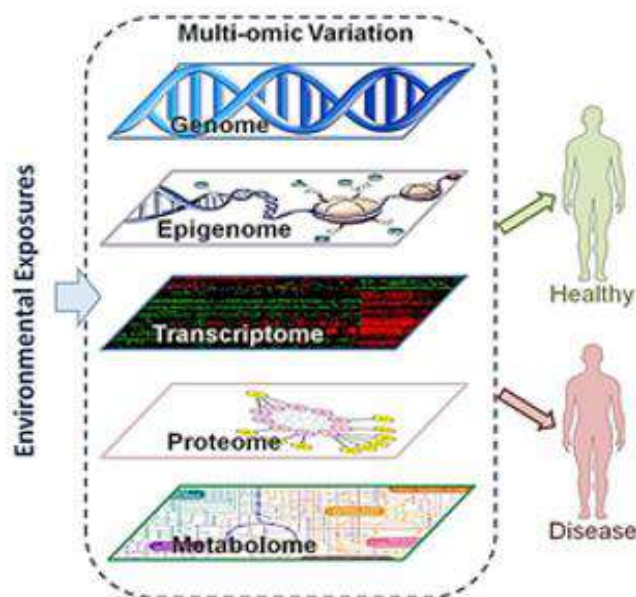


Figura 11: La regulación de las respuestas celulares a los estímulos ambientales puede darse a distintos niveles, desde la expresión de génica, la traducción de transcritos o la actividad de las proteínas. Las alteraciones en los mecanismos de regulación del organismo se han asociado a múltiples enfermedades, y nos han permitido profundizar nuestro conocimiento al respecto de estas. Imagen extraída de Sun y Hu, 2016 [10].

El análisis del interactoma mediante redes multicapa también ha sido útil para comprender los efectos de fármacos en los complejos procesos que ocurren dentro de una célula, pues integra los distintos niveles de regulación celular y la forma en que estos se ven alterados en respuesta a determinado medicamento [30]. De igual modo, ciertos patrones de expresión génica anormales se han asociado a distintos tipos de cáncer o a enfermedades crónico-degenerativas [31].

El mapeo de redes cerebrales mediante redes temporales ha adquirido gran importancia también. Las redes multicapa permiten representar detalladamente las interacciones sinápticas entre neuronas y su evolución ante estímulos. La gran mayoría de los estudios al respecto se han centrado en *C. elegans*, un nemátodo que ofrece

numerosas facilidades al momento de estudiar su sistema nervioso, pues además de ser un modelo de estudio ampliamente conocido, también posee un número reducido y limitado de neuronas. Se han empleado redes múltiples para analizar la señalización aminérgica y neuro peptídica entre las 302 neuronas de *C. elegans*, asociando una capa distinta para cada neurotransmisor involucrado en la comunicación sináptica [48]. En humanos el enfoque de las redes múltiples se ha empleado para estudiar las relaciones entre distintas zonas del cerebro y las respuestas cerebrales en reposo mediante procedimientos no invasivos, tales como la resonancia magnética [49].

En ecología, el enfoque de las redes multicapa ha sido de suma utilidad para comprender las interacciones entre los distintos organismos que componen un ecosistema. La construcción de redes tróficas no solo ha permitido comprender mejor el flujo de materia y energía dentro de un ecosistema, sino también la importancia de las relaciones simbióticas y cómo estas afectan la estabilidad del ecosistema y las poblaciones [11]. Dentro de las redes tróficas se consideran distintos factores, incluyendo el peso o fuerza de la interacción y el tipo de esta (mutualismo, amensalismo, parasitismo, comensalismo, depredación, etc.) [50]. El análisis mediante redes temporales ha arrojado información sobre el cambio de las interacciones entre organismos y la pérdida de biodiversidad en respuesta al cambio climático, a la deforestación o a la contaminación [50].

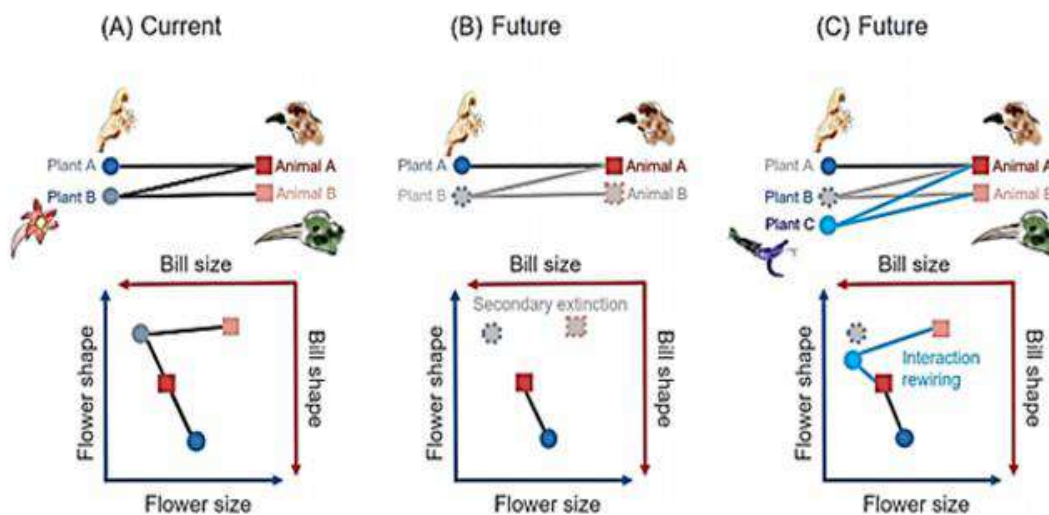


Figura 12: En un ecosistema se dan múltiples interacciones entre distintos organismos. La estabilidad del ecosistema en sí mismo, depende en gran medida de estas interacciones, pues la extinción de una especie puede conllevar a extinciones secundarias de especies que, en mayor o menor medida, dependían de las poblaciones de la primera especie. Imagen extraída de Schleuning et al., 2020 [11].

2.5 Inferencia de redes

Hemos visto que la creación de redes temporales y redes multiplex es posible mediante la delta de Kronecker, dicha función permite conectar los nodos de una capa con sus respectivos nodos réplica en las capas adyacentes. Sin embargo, al intentar construir

las redes de cada capa nos encontramos con un problema: Construir una red a partir de los datos experimentales. Como solución se han propuesto numerosos algoritmos que llevan a cabo la tarea de inferir las posibles asociaciones entre los componentes de una red partiendo de las abundancias de secuencias en las muestras, mismas que pueden asociarse a OTUs o transcritos de genes.

La forma más sencilla de generar una red a partir de una matriz de similitud es eliminando todos aquellos elementos de la matriz cuya similitud sea demasiado débil, es decir, que se encuentre por debajo de determinado umbral [12]. A partir de la matriz resultante se genera una matriz de adyacencia, donde todos aquellos valores por encima del umbral seleccionado son considerados como conexiones entre las secuencias asociadas a algún gen o microorganismo específico, mismos que representamos como nodos dentro de la red [51].

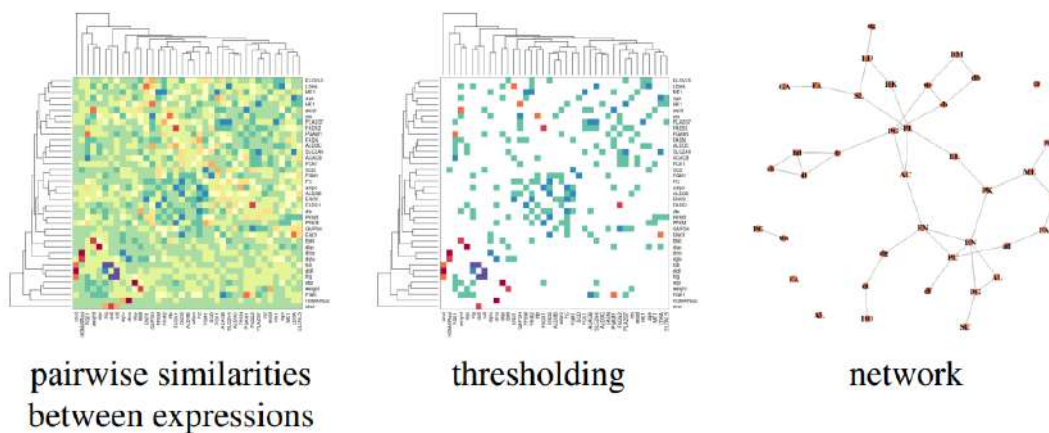


Figura 13: El principio básico de un algoritmo de inferencia de redes es la generación de una matriz de similitud que represente las relaciones de co-abundancia de las secuencias identificadas a partir de los datos muestrales. Cada una de las distintas secuencias encontradas serán consideradas como nodos de la futura red. Todos los valores de la matriz de similitud que se encuentren por debajo del umbral establecido por el investigador, son descartados. A partir de los valores restantes se genera una matriz de adyacencia a partir de la cual se construye una red. Imagen extraída de Villa-Vialaneix-nathalie [12].

Un ejemplo de este método de inferencia es el algoritmo WGCNA (*Weighted Gene Correlation Network Analysis*), empleado en el análisis de correlaciones en redes pesadas de genes [51]. El algoritmo procesa las co-abundancias de los transcritos, expresadas en un *heatmap* como el peso asociado a las aristas, en un dendrograma o árbol. Esto permite clusterizar conjuntos de genes con una expresión similar entre las muestras a partir de las ramas principales del dendrograma [52].

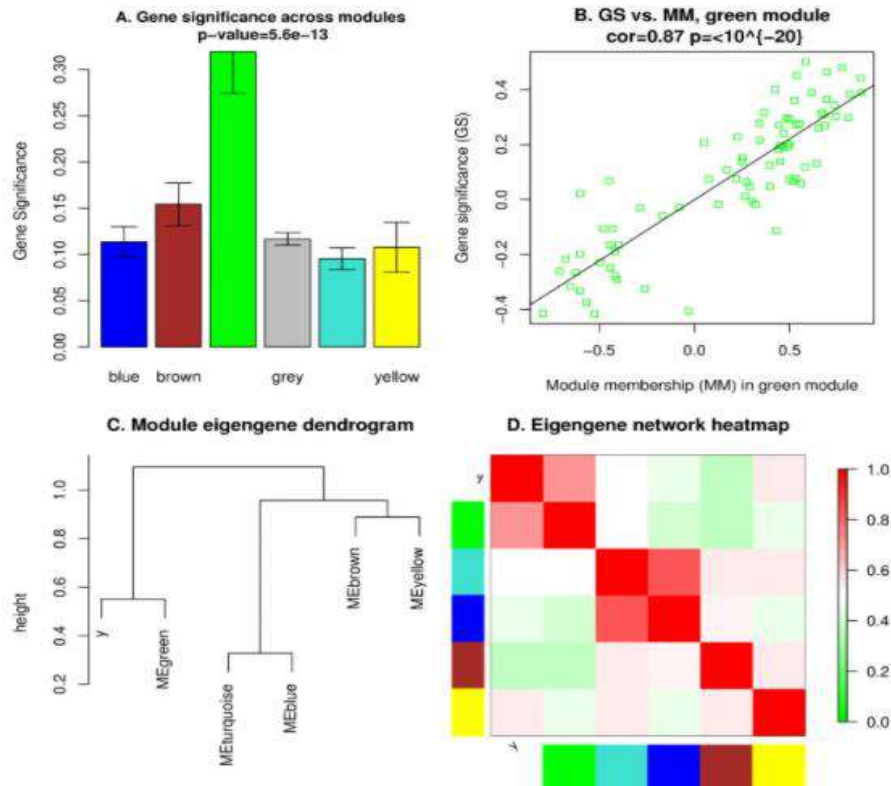
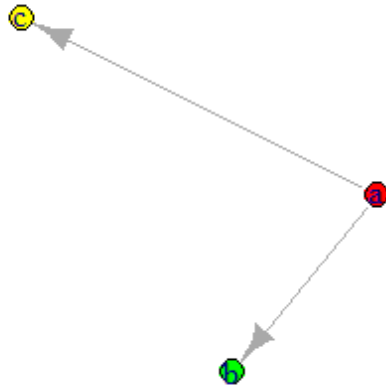


Figura 14: El algoritmo WGCNA construye los módulos a partir de las abundancias de los transcritos en las muestras. Esto permite calcular la significancia media de cada módulo en conjunto con metadatos, como podrían ser categorías GO. Imagen extraída de Langfelder y Horvath, 2008 [13].

Sin embargo, el hecho de que dos elementos de la matriz de similitud posean una correlación alta no implica necesariamente que ambos estén directamente relacionados. Esto implica que el método anteriormente mencionado puede llevar a falsas interpretaciones en cuanto a cómo los nodos están conectados entre sí [51]. Por ejemplo, si tenemos tres genes codificantes para tres productos (a , b y c) que hemos analizado a lo largo de un centenar de muestras, y suponemos que el producto de a regula directamente a b y a c , esperamos que exista una fuerte correlación entre la expresión de estos genes [12]. Sin embargo, cuando calculamos la correlación entre los genes b y c , obtenemos un resultado muy parecido al de los casos anteriores, teniendo estos una correlación fuerte a pesar de no estar directamente vinculados, pero sí relacionados indirectamente mediante su mutua regulación a través del producto del gen a .

```
library(igraph)

#Gráfico de ejemplo
gA <- graph(c("a", "b", "a", "c"))
plot(gA, vertex.color=c("red", "green", "yellow"))
```



```

#Generamos un set de datos de prueba y calculamos la correlación entre los
# nodos.
set.seed(2812)
a <- rnorm(100, 0, 0.1)
b <- 2*a-2+rnorm(100,0,0.1)
cor(a,b)

## [1] 0.9047034

c <- 2*a+4+rnorm(100,0,0.1)
cor(a,c)

## [1] 0.9056657

cor(c,b)

## [1] 0.7763894

```

Figura 15: Como podemos observar en el ejemplo, el producto del gen a regula directamente a los productos de los genes b y c , sin embargo, esto también provoca que exista una fuerte correlación entre los genes b y c a pesar de que la relación entre ambos sea en realidad indirecta.

Claramente, este tipo de correlaciones indirectas puede llevar a interpretaciones erróneas de los datos experimentales, por lo cual es necesario disponer de métodos que permitan descartar este tipo de falsos positivos. Bajo esta perspectiva, los modelos gráficos gaussianos [53], empleados para modelar las relaciones estadísticas entre variables de interés, son una herramienta útil para identificar y descartar interacciones indirectas gracias a que permiten estimar la dependencia condicional entre los nodos [12].

En el siguiente ejemplo, basado en la simulación realizada previamente, podemos observar, además de que las correlaciones tienden a ser menos fuertes que en el primer modelo, que la correlación entre b y c posee valores negativos o significativamente más pequeños con respecto a los valores obtenidos para las correlaciones entre a y los nodos b y c . El modelo calcula las correlaciones parciales, es decir, la correlación entre los residuos de los modelos lineales generados a partir de las variables [12].

```

#Correlaciones mediante un modelo lineal
cor(lm(a~c)$residuals,lm(b~c)$residuals)

## [1] 0.7542552

cor(lm(a~b)$residuals,lm(c~b)$residuals)

## [1] 0.756993

cor(lm(b~a)$residuals,lm(c~a)$residuals)

## [1] -0.2378752

```

Es a partir de este principio que actúan algoritmos de inferencia de redes tales como ARACNE, CLR y MRNET. Estos algoritmos, además, se combinan con estimadores de entropía con el fin de corregir los sesgos asociados al ruido y ajustar la varianza de los datos [12]. Las redes generadas a partir de ellos se conocen como redes de información mutua, que dependen de una correlación fuerte entre dos variables a partir de modelos de dependencia no-lineal, que con frecuencia se ajustan mejor a las nubes de datos que los modelos lineales. Por consiguiente, los algoritmos premian las interacciones directas sobre las indirectas, otorgando valores bajos a las correlaciones entre nodos no relacionados directamente [14]. Cabe destacar que si bien, los algoritmos previamente mencionados fueron diseñados para inferir redes de co-expresión de genes, al emplear el principio de co-abundancias para generar dichas redes también permite su potencial uso en la generación de redes ecológicas bacterianas. El algoritmo ARACNE, por ejemplo, elimina las conexiones más débiles por cada triplete de aristas [54].

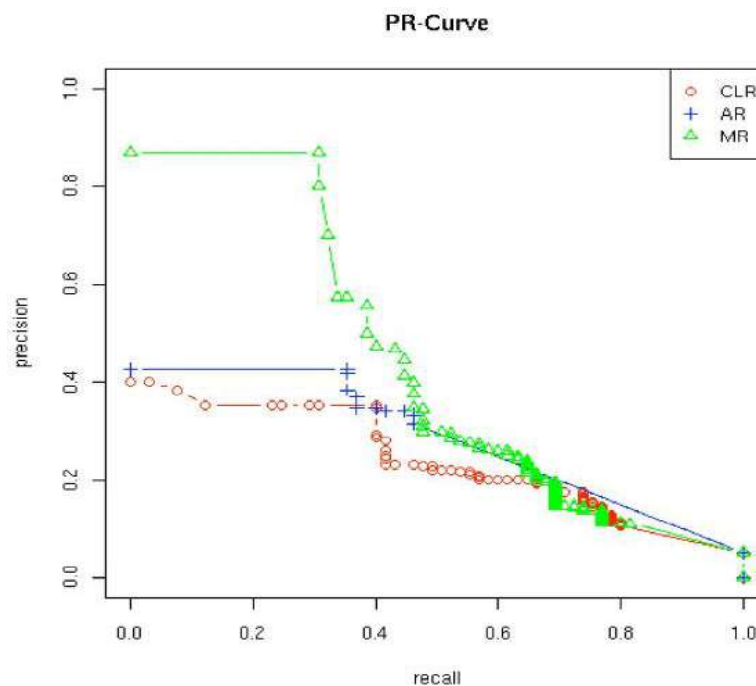


Figura 16: Un gráfico de curvas PR (Precision-Recall curves) que relaciona la precisión de cada algoritmo (CLR, ARACNE y MRNET), es decir, el valor predictivo de los modelos $[TP/(TP + FP)]$ con la recuperación

o sensibilidad del modelo $[TP/(TP + FN)]$. Podemos observar que el poder predictivo de los modelos disminuye con el aumento de la sensibilidad, en otras palabras, la capacidad del modelo de predecir los valores de la nube de datos disminuye conforme este se vuelve más selectivo respecto a aquellos valores que el algoritmo considera pertinente medir, al tiempo que descarta aquellos que no. Imagen extraída de Meyer et al., 2008 [14].

Se han planteado numerosos métodos para la inferencia de redes ecológicas a partir bases de datos de microbiota, estos incluyen algoritmos de inferencia de matrices de similitud y disimilitud o distancias; por lo general empleando los coeficientes de correlación de Pearson o de Spearman [55]. Ambos tipos de matrices son opuestos, pues mientras las primeras adquieren valores más altos entre mayor sea la correlación entre dos variables, las segundas lo hacen conforme aumenta la distancia entre las mismas. Sin embargo, por sí solos estos métodos carecen de la capacidad de diferenciar la presencia de correlaciones falsas, además de ser poco sensibles ante bajos números de muestra [56].

Debido a tales problemáticas se han planteado diversos modelos que buscan enfrentar las limitaciones de los primeros algoritmos. Podemos destacar el algoritmo SparCC (*Sparse Correlations for Compositional data*), mismo que calcula las correlaciones lineales de Pearson a partir de haber aplicado previamente una transformación logarítmica a los datos, esto con el propósito de normalizarlos y reducir así la oblicuidad que pueda presentar su distribución [15].

$$y_{ij} = \log \frac{x_i}{x_j} = \log(x_i) - \log(x_j)$$

El algoritmo asume previamente que el número de componentes, ya sean estos OTUs o genes, es alto; y que la mayoría de los componentes de la red no estarán fuertemente correlacionados, aunque el algoritmo es particularmente robusto a este último supuesto. A pesar de emplear la transformación logarítmica, el algoritmo no asume que los datos tengan estrictamente una distribución normal, sino que el uso de dicha transformación está motivado sencillamente por la facilidad que implica su implementación al procesar los datos. Esto tiene la ventaja de que el valor y_{ij} contiene información de las abundancias de los componentes, al ser una razón entre cada par de nodos, además de que esta relación es independiente del resto de OTUs de la base de datos. Cabe destacar que las correlaciones en este algoritmo pueden adoptar cualquier valor real y no se limitan a valores positivos (pues considera también correlaciones negativas potencialmente asociadas a relaciones antagónicas), razón por la cual se considera un umbral para seleccionar las conexiones entre nodos [15]. Tanto la correlación de Pearson como la de Spearman asumen la linealidad de los datos, por lo que si la correlación entre dos nodos es no-lineal, el algoritmo asigna una correlación baja.

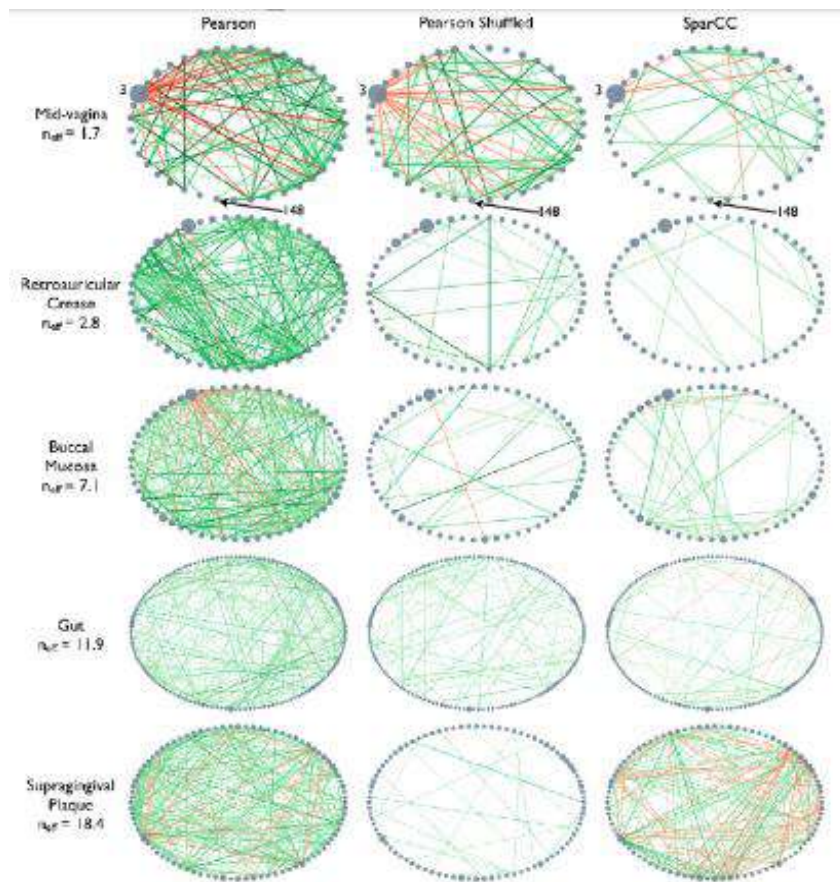


Figura 17: Empleando datos de abundancias del 16S-rRNA extraídos del HMP, se construyeron redes empleando la correlación de Pearson, a la izquierda, y el algoritmo SparCC, a la derecha. También se emplearon datos estocásticos para una tercera red, en el centro. En verde se muestran las correlaciones positivas y en rojo las correlaciones negativas. El hecho de que la correlación de Pearson haya inferido conexiones a partir de los datos simulados indica que algunas de las conexiones podrían ser realmente indirectas. Imagen extraída de Friedman y Alm, 2012 [15].

Otra metodología que ha sido empleada para la inferencia de redes ecológicas es SPIEC-EASI (*SParse Inverse Covariance Estimation for Ecological Association Inference*), que realiza dicha tarea aplicando primero una transformación de los datos originales para luego estimar el gráfico de la red seleccionándolos por vecindad (*MB*) o por medio de la covarianza inversa (*glasso*), permitiendo en ambos casos seleccionar las conexiones que se incluirán en la red [16]. A diferencia de SparCC, que emplea como base la covarianza para estimar las conexiones, SPIEC-EASI se fundamenta en la independencia condicional, es decir, si en un par de nodos, ninguno de estos proporciona información adicional sobre el estado del otro, se les considera condicionalmente independientes. El resultado es una red no-dirigida que considera únicamente correlaciones positivas entre los OTUs [16].

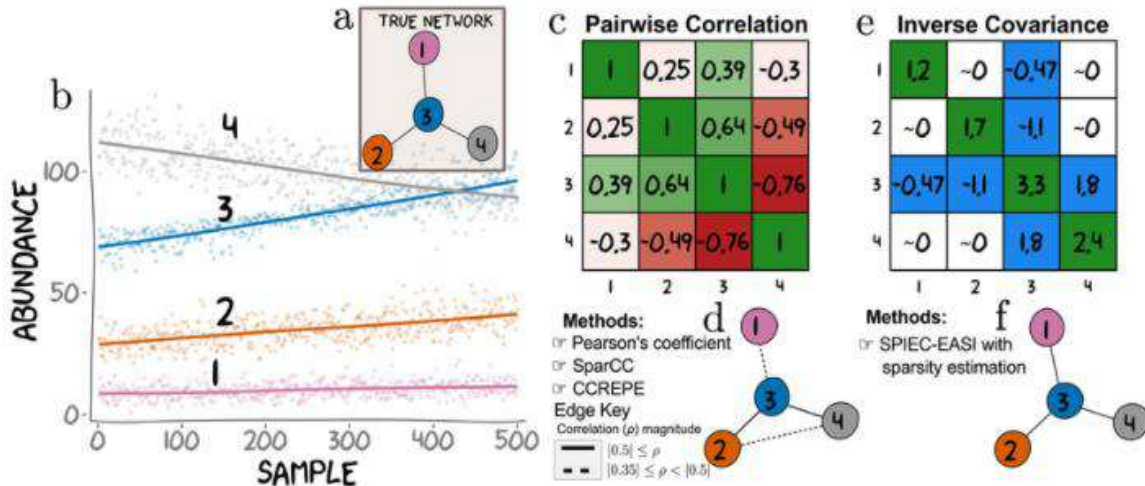


Figura 18: En la imagen se simuló una base de datos con distribución binomial (b) de acuerdo con una red prediseñada (a), para luego emplear distintos algoritmos para inferir dicha red a partir de los datos simulados (c, d), probando así la fiabilidad de cada uno de estos algoritmos en contraste con el método SPIEC-EASI (e, f), siendo SPIEC-EASI el algoritmo con mayor capacidad para eliminar conexiones indirectas. Imagen extraída de Kurtz et al. 2015 [16].

Si bien los algoritmos de coocurrencia se emplean a menudo para identificar interacciones entre especies, estos no son capaces de inferir *per se*, asociaciones ecológicas [16]. Se considera que estos métodos son una herramienta útil para inferir potenciales interacciones entre taxones valiéndose del fundamento de que las poblaciones de cualquier organismo son dependientes de la abundancia de otros organismos, sin embargo, estas interacciones no suelen darse entre pocos organismos ni en condiciones controladas como sí ocurriría dentro de un laboratorio, sino que en un ecosistema microbiano podemos encontrar miles de taxones interactuando de distintas formas bajo condiciones variables [17]. Si bien, se consideran herramientas útiles para inferir posibles asociaciones, no pueden ser consideradas como metodologías infalibles, aunado al hecho de que la aplicación de distintos algoritmos a distintas bases de datos suele presentar importantes discrepancias en los resultados que arrojan; razón por la cual deben ser entendidas como lo que son, herramientas de inferencia [57]. Para establecer una conexión entre los componentes de una red es importante obtener información extra que permita conocer la veracidad de la correlación [17].

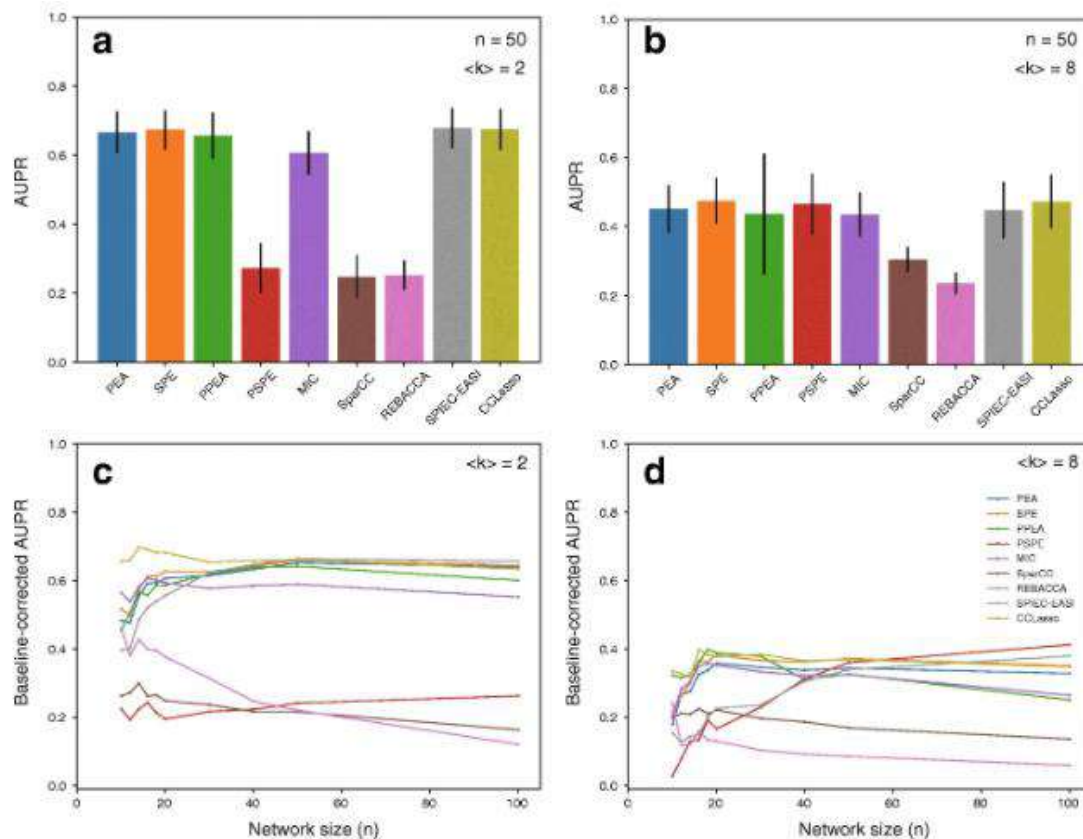


Figura 19: Análisis de las discrepancias en los resultados obtenidos mediante distintos algoritmos de inferencia para una misma base de datos. Se consideró para ello una red con un total de 50 componentes y dos valores para el degree promedio ($\langle k \rangle = 2m/n$), donde n es el número de nodos y m el número de conexiones obtenidas. Se observa que los distintos algoritmos (a, b) presentan visibles diferencias entre los valores AUPR obtenidos (área bajo la curva PR, misma que relaciona la precisión de cada algoritmo con su sensibilidad). También podemos observar que algunos algoritmos son muy sensibles al tamaño de la red, y que a pesar de aumentar este parámetro las discrepancias entre los algoritmos no necesariamente disminuyen y, por el contrario, aumentan. Imagen extraída de Hirano y Takemoto, 2019 [17].

2.6 La microbiota intestinal y su impacto en la salud humana

La microbiota puede definirse como el conjunto de microorganismos (bacterias, hongos, arqueas, protozoos y virus) que proliferan y coexisten en los tejidos de un organismo multicelular. Los humanos tenemos una microbiota asociada a tejidos superficiales y cavidades; incluyendo el tracto digestivo, el tracto respiratorio, el tracto urinario, la piel, entre otras [58]. Los microorganismos que componen la microbiota poseen complejas relaciones con el huésped, a menudo siendo comensalistas o mutualistas que desempeñan un papel importante para el buen funcionamiento del organismo. Estas interacciones son dinámicas y al verse alteradas pueden tener un impacto sustancial en la salud del individuo. Diversos estudios ómicos y fisiológicos (tanto en modelos animales como en estudios en humanos) han revelado que la constitución de las comunidades microbianas tiene una influencia significativa como

mediadores en el desarrollo de enfermedades en respuesta a estímulos ambientales [18]. Las alteraciones en respuesta a una alimentación no balanceada y escaso ejercicio pueden provocar que los microorganismos con relaciones simbióticas con el huésped adquieran características patogénicas en respuesta a determinado estímulo, desencadenando el desarrollo de una enfermedad. La microbiota ha probado tener un papel importante en la digestión, la regulación de la respuesta inmune, la función endocrina del intestino, la inactivación de toxinas, el metabolismo de fármacos y nutrientes, y la señalización neuronal, llegando a impactar incluso el estado emocional [59].

Dada la inmensa diversidad de estilos de vida, patrones genéticos y epigenéticos, la gran variedad de taxones microbianos, y la complejidad inherente a la relación entre los tejidos y la microbiota asociada, el límite entre una microbiota saludable de una que no lo es, resulta difuso. A pesar de ello, se toma como supuesto fundamental la composición y funcionamiento de la microbiota de individuos considerados sanos [60]. Para ello se toman como base las abundancias relativas de los microorganismos de interés, en lo concerniente a este documento, de las bacterias. Además, cabe destacar que el tipo de interacciones que presenta cierto grupo de microorganismos puede ser muy distinto de una especie a otra de una misma familia, o incluso de una cepa a otra de la misma especie, por lo que hacer generalizaciones al respecto puede resultar con frecuencia, contraproducente [61]. Por lo general, se considera que una microbiota más diversa está a menudo asociada a individuos sanos, pero la diversidad está también fuertemente ligada a la velocidad de tránsito intestinal, por lo que la diversidad y riqueza no pueden ser considerados como indicadores absolutos. Por ejemplo, un tránsito prolongado de la materia fecal puede contribuir a una mayor diversidad bacteriana, pero no necesariamente a una más saludable [60].

Adquirimos nuestra microbiota durante el parto, cuando nuestro organismo es colonizado por millones de microorganismos distintos. A partir de este momento, la microbiota cambia gradualmente a lo largo de nuestra vida en respuesta a diversos estímulos que van desde el tipo de alimentación, la actividad física, la exposición a determinadas sustancias o la edad. El envejecimiento, por ejemplo, se ha asociado a una reducción en la diversidad microbiana en respuesta a cambios en la actividad del sistema inmune [62]. Los microorganismos que componen la microbiota son los responsables de sintetizar metabolitos secundarios que tienen un impacto a distintos niveles en la fisiología del organismo, tales como la producción de ácidos grasos de cadena corta (AGCC), la síntesis de lipopolisacáridos (LPS), o la biosíntesis de vitaminas y aminoácidos esenciales [63]. Los datos metagenómicos deben apoyarse también en el entendimiento de los mecanismos que conectan los fenómenos observados con la diversidad microbiana [64].

Cabe destacar, también, que el uso masivo de antibióticos alrededor del mundo ha modificado sustancialmente los patrones de composición de la microbiota humana, y dado que casi la totalidad de humanos han estado expuestos a estos fármacos al menos una vez en su vida, se considera que esto da lugar a un sesgo en la percepción de lo que consideramos como una microbiota *core* o central [65]. De igual modo, los patrones dietéticos han sufrido importantes cambios en la mayoría de las sociedades desde la

introducción de los alimentos pre-procesados, así como también modificaciones en el estilo de vida y la actividad física, el uso de probióticos y prebióticos, e intervenciones más radicales como la del uso de la fagoterapia o el trasplante de microbiota [66]. Todos estos factores, en mayor o menor medida, constituyen un sesgo en la percepción del funcionamiento de la microbiota, más aún, considerando que la implementación de las tecnologías que nos permitieron aunar en la complejidad de tales sistemas es posterior a estos eventos. En los países desarrollados, donde muchos de estos fenómenos se han presentado con una mayor extensión en la población, se ha observado una reducción gradual de la diversidad microbiana. Esta reducción incluye géneros como *Bacteroides*, *Prevotella*, *Desulfovibrio*, *Lactobacillus* y *Oxalobacter*, que ha su vez se ha asociado a una mayor prevalencia de enfermedades crónicas en contraposición con un aumento en los estándares de vivienda, seguridad alimentaria, acceso a servicios sanitarios e higiene [67]. Dentro de esto podemos destacar la hipótesis de la higiene, misma que estipula que la exposición limitada a distintos antígenos relativamente comunes durante la infancia, ha dado paso a un desarrollo deficiente del sistema inmune, provocando que este reaccione de manera anómala ante estímulos ambientales cotidianos, siendo este fenómeno particularmente asociado a la aparición de alergias y enfermedades crónico-inflamatorias. Esta respuesta inmune anómala también conlleva alteraciones en la composición de la microbiota, alteraciones que permiten retroalimentar el estado crónico-inflamatorio de los individuos afectados [68].

El mantenimiento de la homeostasis depende de múltiples factores tanto ambientales como genéticos que, en su conjunto, afectan la composición de la microbiota. La disbiosis, como resultado de la alteración de las poblaciones microbianas, conduce a un estado de enfermedad. Una dieta rica en fibra y baja en grasas y proteínas animales conduce a la proliferación de bacterias fermentadoras de polisacáridos, mismos que son degradados a una variedad de compuestos que, entre otras cosas, promueven la producción de moco y la función de barrera del epitelio intestinal [18]. A partir de esta fermentación se obtienen ácidos grasos de cadena corta (AGCC), tales como el butirato, el propionato y el acetato; además del aumento en la producción de metano y la disminución del pH. Los AGCCs, además de ser una fuente de energía adicional para los colonocitos, interactúan con los receptores acoplados a proteínas G de las células L entero-endócrinas: GPCR-41 y GPCR-43, que a su vez inducen la secreción de péptido YY (PYY) y de péptido-1 similar al glucagón (GLP-1), que contribuyen a la sensación de saciedad, a la liberación de insulina y a el metabolismo de la glucosa [69]. Por su parte, el butirato favorece la beta-oxidación y el consumo de oxígeno en el intestino, ayudando a mantener un ambiente anaerobio, además de la regulación de la respuesta inmune mediante la estimulación de los linfocitos T reguladores [70].

Una dieta rica en proteínas y grasas animales, así como baja en fibra promueve un estado de disbiosis, lo que favorece la degradación de la mucosa por parte de las propias bacterias fermentadoras (en respuesta a la escasez de sustrato), la pérdida de la barrera intestinal y la filtración de compuestos de origen microbiano (PAMPs), incluyendo lipopolisacáridos (LPSs), al torrente sanguíneo, provocando un estado de inflamación que se ve agravado ante la baja producción de AGCCs [71]. Esto también contribuye a retrasar la sensación de saciedad y altera el metabolismo de la glucosa al favorecer la

resistencia a la insulina. Un tránsito lento aunado a un consumo rico en grasas y proteínas promueve igualmente un metabolismo proteolítico que conlleva el aumento de aminoácidos aromáticos, ácidos grasos de cadena ramificada, fenoles, indoles, gases y aminas, que contribuyen a aumentar el pH. Aunque, de igual modo, ha de notarse que la producción de indol y sus derivados a partir de la fermentación bacteriana de fibra favorece el metabolismo de la glucosa [72].

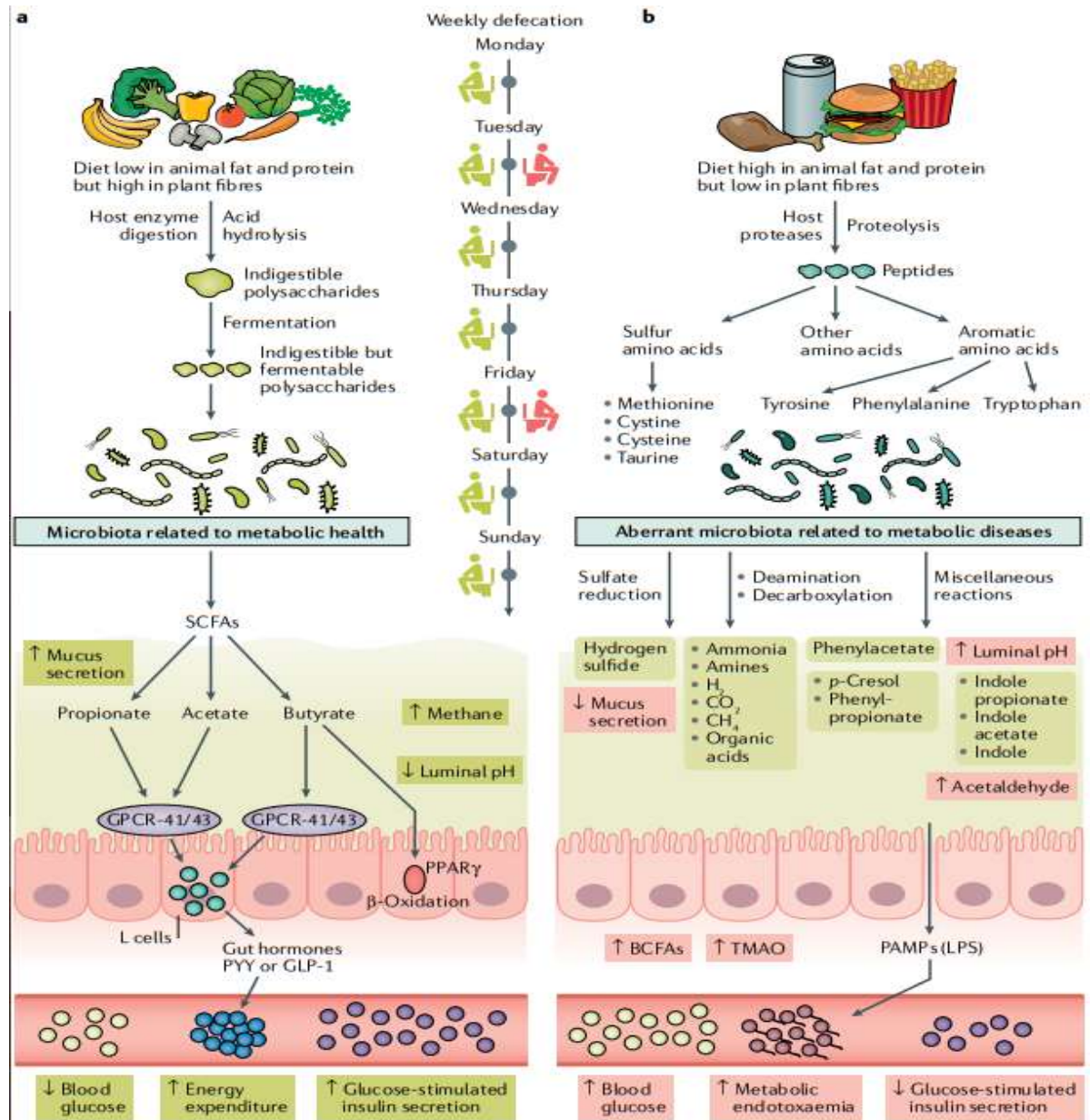


Figura 20: Una dieta rica en fibra vegetal favorece la proliferación de bacterias productoras de AGCCs que estimulan la producción de moco y metano, además de tener propiedades antiinflamatorias. Por el contrario, una dieta rica en proteínas y grasas animales altera la producción y captación de insulina, además de promover un estado de inflamación crónica [18]. Imagen extraída de Fan y Pedersen, 2021 [73].

La microbiota intestinal sufre cambios en individuos que sufren obesidad con respecto a individuos con IMC normal. Se ha asociado a algunas especies productoras de ácidos grasos de cadena corta como *Eubacterium ventriosum* y *Roseburia intestinalis* con la obesidad [74]. Sin embargo, la síntesis de algunos tipos de AGCCs como el butirato, el propionato y el acetato por bacterias fermentadoras de fibra como *Oscillospira* y *Akkermansia muciniphila*, o arqueas como *Methanobrevibacter smithii*, se han asociado al efecto contrario [75]. De igual modo, se ha asociado la presencia de bacterias fermentadoras de glutamato como *Bacteroides thetaiotaomicron* con la delgadez, donde su abundancia es inversamente proporcional a la concentración sérica de glutamato [76]. Además, la obesidad también se ha asociado con una reducción en la conjugación unidireccional, es decir, la transferencia de material genético entre bacterias, además de un aumento del estrés oxidativo derivado de una reducción de la actividad de las superóxido reductasas bacterianas, efectos que parecen ser reversibles mediante trasplantes de materia fecal en ratones [77].

En individuos que padecen diabetes tipo II y prediabetes existe evidencia de una microbiota alterada con respecto a individuos sanos. Si bien esto no supone necesariamente que dicha alteración sea responsable de la resistencia a la insulina, la evidencia indica que la variación de las abundancias relativas podría tener un efecto indirecto sobre la secreción de insulina. Estudios en roedores indican que la hiperglucemia puede aumentar la permeabilidad de la barrera intestinal mediante la alteración de la actividad de GLUT2 en células epiteliales y el debilitamiento de las uniones estrechas, provocando una filtración de compuestos de origen bacteriano al torrente sanguíneo, lo que deriva en un estado de inflamación crónica [78]. En individuos con prediabetes se ha identificado una considerable reducción en las abundancias relativas de taxones asociados a la síntesis de butirato, especialmente en el caso de *Akkermansia muciniphila*, además de un aumento en poblaciones asociadas a la producción de factores proinflamatorios [79].

Cabe aclarar que el tratamiento farmacológico de la diabetes tipo II ha hecho difícil el estudio del papel de la microbiota en este padecimiento, debido a los factores de confusión derivados de los efectos que dichos fármacos puedan tener sobre las poblaciones microbianas. Por ejemplo, se ha documentado que la metformina altera las abundancias de diversos géneros bacterianos como *Escherichia* e *Intestinibacter*, aunque parece tener un impacto positivo en la producción de propionato y butirato, así como de ácidos biliares no-conjugados, induciendo así la gluconeogénesis intestinal y la consecuente reducción de la glucemia [80]. Sin embargo, experimentos en roedores libres de gérmenes han indicado que los efectos de la metformina permanecen inalterados a pesar de la ausencia de una microbiota, lo que podría indicar que los efectos de dicho fármaco sobre la microbiota podrían ser más bien modestos [81]. A pesar de la evidencia inicial, los cambios en la microbiota observados en individuos con diabetes tipo II parecen no ser específicos de dicha enfermedad, pues se han encontrado patrones similares también para el caso de otros trastornos crónicos caracterizados por una inflamación leve o clínicamente silenciosa. Además de que no se ha logrado reproducir con éxito el fenotipo de la diabetes al trasplantar materia fecal de ratones enfermos a ratones libres de gérmenes [79].

Al desbalance o cambio en las abundancias relativas normales de los taxones bacterianos de la microbiota intestinal, se le conoce como disbiosis microbiana intestinal. La disbiosis se ha relacionado a un gran número de enfermedades cardio-metabólicas, siendo una de las más frecuentes la arteriosclerosis [82]. Los individuos que padecen arteriosclerosis a menudo presentan concentraciones altas de glucosa sérica, insulina y lípidos, además de inflamación crónica de bajo grado y resistencia a la insulina, por lo que suele presentarse en personas que también padecen diabetes tipo II. Además, en casos más severos la arteriosclerosis también puede ir acompañada por cardiopatía isquémica, donde también se ha documentado los patrones de abundancias microbianas asociados a la disbiosis. Sin embargo, como igual sucede en el caso de la diabetes, el hecho de que las personas con arteriosclerosis estén fuertemente medicadas dificulta conocer cuál es el papel de la microbiota en el desarrollo de la enfermedad [83].

Estudios realizados sobre muestras fecales de individuos con enfermedades cardio-metabólicas parecen exhibir abundancias enriquecidas de bacterias de la familia Enterobacteriaceae, tales como *Escherichia coli*, *Klebsiella* spp. y *Enterobacter aerogenes*; en contraste con una disminución de *Bacteroides* spp. y *Faecalibacterium prausnitzii*. El cambio en dichas abundancias se ha correlacionado a un aumento en la síntesis de lipopolisacáridos, el transporte de aminoácidos, además de la síntesis de trimetilamina (TMA) y aminoácidos aromáticos como el triptófano. Por el contrario, existe una correlación inversa de la actividad de los AGCCs, especialmente de la producción del butirato, así como del metabolismo de vitaminas. En términos generales podemos aseverar que el microbioma se torna menos fermentativo al tiempo que promueve un estado de inflamación crónica leve [84]. Recientemente también se ha asociado una mayor abundancia de *Ruminococcus*, *Acinetobacter* y *Veillonella* spp., así como una disminución de *Alistipes*, *Faecalibacterium* y *Oscillibacter* spp. como potenciales marcadores de la insuficiencia cardíaca isquémica [85].

La N-óxido de trimetilamina (TMAO) se ha identificado como un factor de especial relevancia en el desarrollo de enfermedades vasculares isquémicas. La trimetilamina (TMA) se sintetiza mediante la actividad enzimática bacteriana a partir de la fosfatidilcolina, la lecitina y la l-carnitina, compuestos abundantes en una dieta a base de productos cárnicos. Cuando la TMA ingresa al flujo sanguíneo del sistema porta enterohepático, viaja al hígado donde es oxidada a TMAO [86]. Experimentos con roedores han mostrado que altas concentraciones séricas de TMAO provocan una aceleración del desarrollo de arteriosclerosis, aumentando la agregación plaquetaria y el riesgo de trombosis. Estudios con el inhibidor 3,3-dimetil-1-butanol de la síntesis de TMA han sido exitosos en reducir el riesgo de desarrollar arteriosclerosis, trombosis, infarto al miocardio, isquemia arterial y accidente cerebrovascular [87].

Existen inconsistencias en cuanto a los resultados, ya que experimentos con roedores alimentados con una dieta enriquecida con l-carnitina, si bien poseen concentraciones elevadas de TMAO como se esperaría, ver reducido al tamaño de las lesiones aórticas características de la isquemia vascular, por lo que la l-carnitina podría tener un efecto benéfico en contraposición a su condición como precursor de la TMA [88]. Sin embargo, estudios recientes en ratones sugieren que estas discrepancias podrían ser el resultado

de una combinación de factores genéticos, la constitución de la microbiota y la dieta, que al interactuar entre sí predisponen al individuo en mayor o menor medida a desarrollar arteriosclerosis; por lo que, si bien la concentración de TMAO puede ser útil como biomarcador, no supone necesariamente el desarrollo consecuente de la patología [18].

La disbiosis intestinal también se ha relacionado al desarrollo de síndrome metabólico hepático, especialmente de hígado graso no-alcohólico (HGNA). El HGNA comprende un amplio abanico de enfermedades, siendo la esclerosis no-alcohólica la más agresiva de sus manifestaciones [88]. Pacientes con HGNA tienden a mostrar abundancias enriquecidas de *Clostridium*, *Anaerobacter*, *Streptococcus*, *Escherichia* y *Lactobacillus*, al tiempo que presentan poblaciones reducidas de *Oscillospira*, *Flavonifaractor*, *Odoribacter* y *Alistipes* spp. Las proteobacterias y enterobacterias parecen tener especial relevancia en el desarrollo de la enfermedad al verse aumentadas sus abundancias relativas. Dicha relación podría ser el resultado de compuestos hepatotóxicos sintetizados por la microbiota alterada, tales como 2-butanona y 4-metil-2-pentanona, se han hallado en concentraciones anormalmente altas en individuos con HGNA [19].

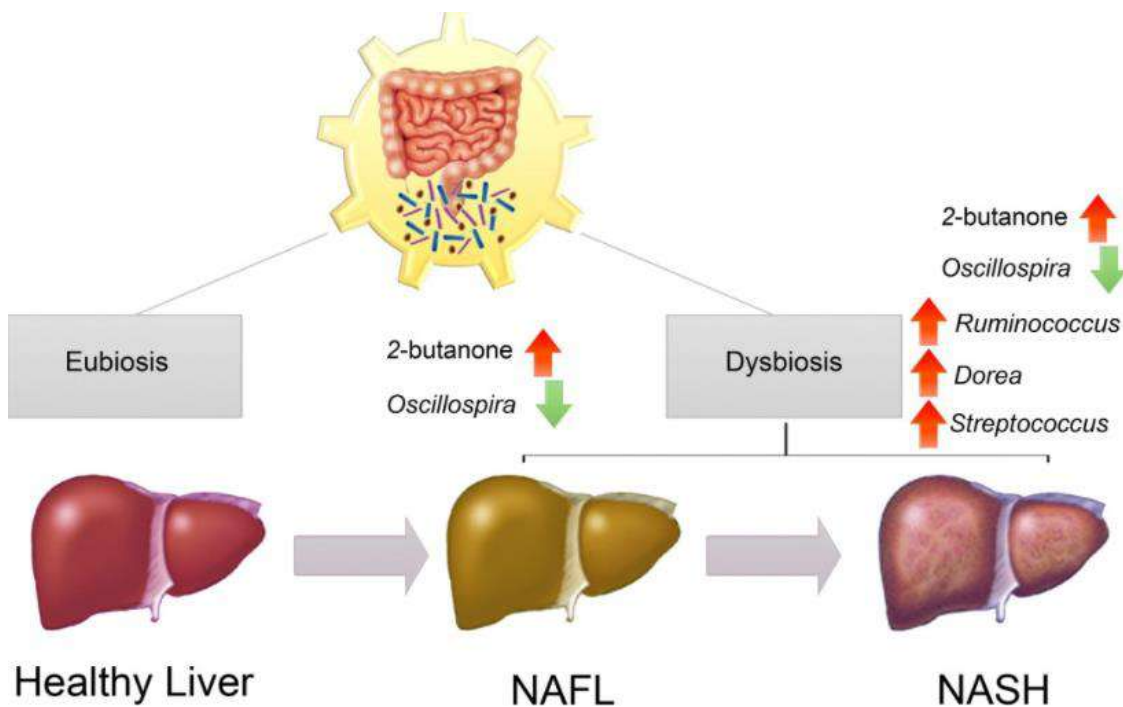


Figura 21: La disbiosis microbiana intestinal se ha asociado al desarrollo de hígado graso no-alcohólico que puede incluso complicarse hasta convertirse en esteatohepatitis, es decir, hasta desarrollar un proceso necroinflamatorio del tejido hepático. La abundancia anormalmente alta de bacterias como *Ruminococcus*, *Dorea* y *Streptococcus*, así como la disminución en las abundancias de *Oscillospira*, se han propuesto como marcadores del desarrollo de este padecimiento, que podría asociarse a una mayor producción de compuestos hepatotóxicos como la 2-butanona por parte de la microbiota aberrante. Extraída de Del Chierico et al., 2017 [19].

Además, la mayor abundancia de bacterias productoras de etanol como *Escherichia coli* contribuye también al desarrollo de la patología. El etanol es capaz de promover la actividad del factor nuclear κ B (NF- κ B), asociado a la respuesta a estrés y a la regulación de la actividad inmune, y cuya activación exacerbada es capaz de provocar daño tisular, contribuir a la endotoxemia en el sistema porta enterohepático y alterar la integridad de la barrera intestinal. Esto provoca un estrés añadido en el tejido hepático, que ve comprometida su capacidad de inactivar toxinas que a su vez contribuyen a retroalimentar el estrés oxidativo sobre los hepatocitos y el daño tisular, generando un estado de inflamación crónica del tejido hepático [86]. Otras bacterias como *Klebsiella pneumoniae* parecen acelerar el desarrollo de la enfermedad, pero aún se requiere de una mayor investigación que ayude a esclarecer los mecanismos moleculares por los cuales la microbiota en general afecta el desarrollo de este tipo de patologías [89].

Estudios sobre la microbiota de niños con desnutrición han arrojado indicios de cambios en la microbiota con respecto a individuos sanos. La microbiota sana en la infancia temprana se caracteriza por la floración temprana de bifidobacterias como *Bifidobacterium longum* y *Bifidobacterium pseudolongum*, responsables del metabolismo de la leche materna, cuyas poblaciones después se reducen en favor de la proliferación de bacterias anaerobias. Infantes con desnutrición aguda muestran una pérdida temprana de bifidobacterias, además de una escasa diversidad bacteriana en comparación con infantes sanos. Se ha señalado que la diversidad bacteriana se puede restablecer mediante una dieta equilibrada y suplementos vitamínicos, aunque no está del todo claro hasta qué punto esto es posible y los efectos sobre la microbiota a largo plazo [20].

A pesar de que la información de la que disponemos actualmente no es suficiente como para comprender los mecanismos que conectan a la microbiota con el desarrollo de ciertas enfermedades, los estudios indican que se trata de una combinación de factores genéticos, ambientales y de la propia constitución de la microbiota, que en su conjunto dan lugar a un gran abanico de fenotipos. La compleja interacción entre las bacterias que componen la microbiota, el huésped y factores externos es aún un terreno desconocido, por lo que se requiere de mucha más investigación para comprender la relación entre tales factores y su papel en el desarrollo de patologías y el mantenimiento de la homeostasis [20].

2.7 Interacción planta-microbiota-insecto

Al igual que los humanos, las plantas poseen una microbiota asociada a sus tejidos con la cual interactúan y guardan una relación simbiótica. Las rizobacterias, por ejemplo, se asocian a las raíces de las plantas y promueven su crecimiento al facilitar la biodisponibilidad de minerales y otros nutrientes. Estas bacterias a menudo presentan una maquinaria enzimática, de la cual carecen las plantas, y que permite fijar elementos como el nitrógeno [90]. Las rizobacterias, además, inhiben el crecimiento de otras bacterias competidoras, muchas de las cuales son patógenas para las plantas. Además de la presencia de rizobacterias, cabe destacar a las llamadas micorrizas, hongos que, en conjunto con dichas bacterias, forman una relación simbiótica con la rizosfera al proveer de agua y algunos nutrientes [91]. A cambio, tanto las micorrizas como las

rizobacterias asociadas a las raíces reciben un suministro constante de azúcares derivadas de la fotosíntesis de la planta. Se ha observado que la explotación del suelo con fines agrícolas altera considerablemente la composición de las poblaciones bacterianas nativas, disminuyendo la diversidad y dificultando la biodisponibilidad de nutrientes en el suelo [92].

Los insectos herbívoros también poseen una microbiota asociada que entra en contacto con las plantas de los que estos insectos se alimentan, a lo cual llamamos eje planta-microbiota-insecto. Debido a que estos insectos perforan los tejidos de las hojas para alimentarse, las plantas están provistas de respuestas bioquímicas que evitan la proliferación de bacterias patógenas vía ácido jasmónico o ácido salicilato como respuesta al estrés, y/o alterando la concentración de compuestos volátiles capaces de atraer parásitos e insectos-plaga [93]. Algunas rizobacterias como *Bacillus amyloliquefaciens* y *Pseudomonas putida*, asociadas a la menta, son capaces de promover la activación de las vías del ácido jasmónico y del ácido salicilato, una actividad que permite a la planta prepararse con antelación ante un eventual ataque de orugas [94]. Un fenómeno similar se ha observado en el caso de bacterias como *Bradyrhizobium japonicum* y *Delftia acidovorans* ante la presencia de bacterias, una actividad que podría estar mediada por señales bioquímicas de plantas adyacentes y que se ve reforzada por dichas bacterias [95]. Se ha observado que bacterias como *Bacillus amyloliquefaciens* están involucradas en la alteración de la síntesis de compuestos volátiles que reducen la probabilidad de atraer a tijeretas (*Doru luteipes*) [95]. Diversas especies de *Pseudomonas* spp. se han asociado a la reducción de la supervivencia de la mosca blanca, así como inhibidoras del crecimiento de áfidos [96].

El cambio climático parece estar teniendo un efecto significativo en la interacción del eje planta-microbiota-insecto y, por ende, en la salud de las plantas. Si bien se considera que mayores concentraciones de CO_2 promueven mayores tasas de crecimiento de la materia vegetal, el aumento repentino de las concentraciones de carbono también altera el crecimiento de las poblaciones de plagas y parásitos, y las tasas de crecimiento de las poblaciones de bacterias patógenas. Esto viene aunado a una reducción de la nutrición de las plantas y una alteración de su fisiología, lo que las deja expuestas a potenciales infecciones con mayor facilidad [91]. Además, el aumento de las concentraciones de ozono a nivel del suelo por la actividad humana es un factor de estrés para las plantas, lo que dificulta una eficaz respuesta al estrés y su tasa de crecimiento. Los cambios en la concentración de CO_2 y O_3 también se han asociado a cambios en la composición de la microbiota del suelo, lo que deriva en una alteración de la biodisponibilidad de nutrientes para las raíces y una mayor exposición a bacterias patógenas [90].

2.8 Uso de redes en la exploración de la microbiota

La interacción entre las bacterias que componen la microbiota es aún en gran medida desconocida. Las poblaciones bacterianas responden a distintos factores que afectan la abundancia de estas bajo condiciones determinadas y en momentos determinados. Si bien, pares de bacterias muestran interacciones que pueden ser simbióticas o antagonicas en condiciones experimentales, fuera de un medio controlado estas

interacciones se pueden ver alteradas por las interacciones con otras bacterias, con el tejido del huésped y por los nutrientes derivados de la dieta del propio huésped y/u otros factores externos [56]. Además, podemos incluir la presencia de otras sustancias, tales como fármacos, otros organismos como parásitos, arqueas o bacteriófagos, y un gran número de variables, en su mayoría desconocidas, y que, como resultado de su efecto, afectan a las abundancias de las poblaciones [61]. Debido a la cada vez mayor evidencia sobre la importancia de la microbiota en el mantenimiento de la homeostasis, la comprensión de las relaciones inter-microbianas se ha tornado cada vez más relevante en el proceso de comprensión sobre el desarrollo de múltiples enfermedades [59].

El uso de redes como una de múltiples herramientas bioinformáticas en la exploración de la microbiota, contribuye a abordar la gran cantidad de datos obtenidos mediante la secuenciación y facilita la comparación de múltiples bases de datos en busca de patrones que puedan servir de punto de partida para comprender más a fondo las interacciones bacterianas [97]. La importancia funcional de las comunidades dentro de las redes es aún difusa, sin embargo, la evidencia parece mostrar que comunidades o *clusters* estrechamente interconectados son comunes en las redes microbianas, lo que sugiere que podrían tener una importancia biológica. Redes generadas mediante simulaciones por métodos bioinformáticos tienden a perder los *clusters* típicamente encontrados con datos reales [98]. Dentro de estas comunidades los nodos con mayor *degree*, que en teoría podrían corresponder a bacterias de especial relevancia dentro de la ecología de la propia microbiota, parecen estar mucho más interconectados que en el resto de la red. Además, también existe evidencia de que estas comunidades están relativamente conservadas entre individuos, aunque cabe destacar que distintos estudios también muestran discrepancias considerables en lo concerniente a este punto [99].

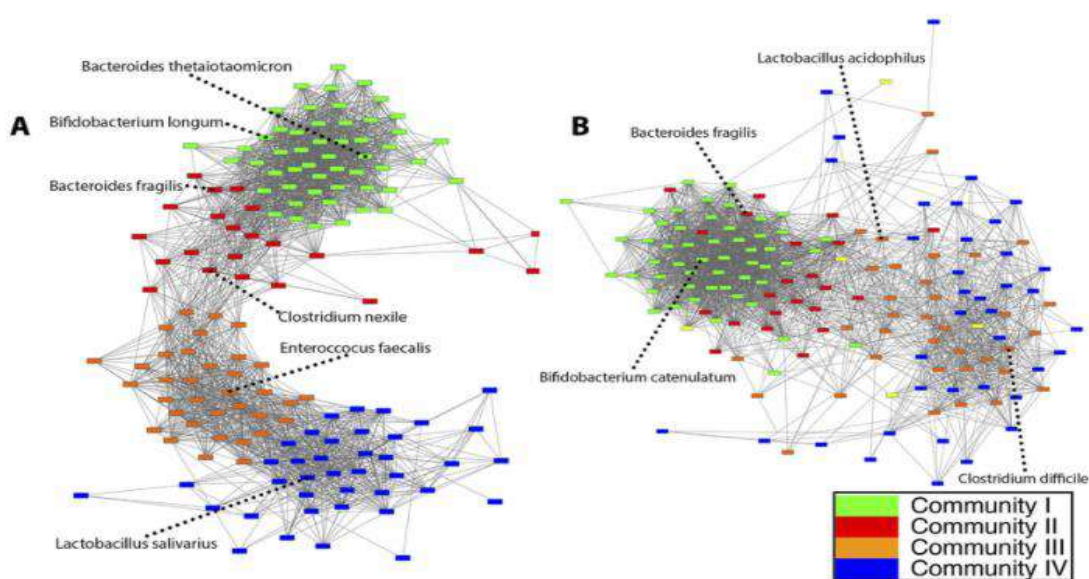


Figura 22: Redes de microbiota inferidas mediante el algoritmo WGCNA. A la derecha se encuentra la red inferida a partir de datos de individuos sanos, y a la izquierda la red correspondiente a pacientes con síndrome de intestino irritable. Mientras que en las comunidades de individuos sanos abundaban géneros como

Bacteroides, *Bifidobacterium*, *Prevotella*, y *Ruminococcus*, asociados al metabolismo fibrolítico y a la síntesis de compuestos antiinflamatorios; en el segundo caso abundaban géneros como *Bacteroides* y *Clostridium*, este último género está asociado al desarrollo de gastroenteritis. Imagen modificada de Smith et al., 2013 [99].

A pesar de que las redes temporales y multiplex son herramientas relativamente recientes en la exploración de la microbiota, han probado ser de gran utilidad para comparar distintas redes. Como se ha mencionado previamente, un conjunto de bacterias no necesariamente interactúa de la misma forma bajo dos o más condiciones o periodos de tiempo distintos, por lo cual conocer en qué medida cambian las interacciones entre los nodos entre capas podría ayudar a comprender cómo la microbiota afecta y se ve afectada por los cambios ocurridos en su medio [98]. Las comunidades bacterianas no suelen permanecer estáticas ante variaciones en las condiciones de su medio, por ejemplo, durante el desarrollo de una enfermedad, en el proceso de envejecimiento, tras un cambio en la dieta o posterior a la ingesta de un fármaco. Debido a esto, algunas comunidades bacterianas tienden a ganar relevancia mientras otras la pierden, evento que tiene un impacto en la arquitectura de la red y que potencialmente puede ayudar a inferir la relevancia de las comunidades a nivel biológico [97].

El uso de redes multicapa emerge como un nuevo paradigma de la teoría de redes, como una herramienta muy útil empleada en multitud de campos, incluida la biología. Este enfoque permite comparar redes heterogéneas y analizar las diferencias entre las mismas, la conservación de módulos entre capas y la identificación de diferencias en las propiedades de los nodos entre capas. Si bien posee numerosas limitaciones dentro del marco de sus posibilidades, es una herramienta útil que puede ayudar a reconocer patrones que podrían potencialmente tener relevancia en la dinámica del sistema biológico [21].

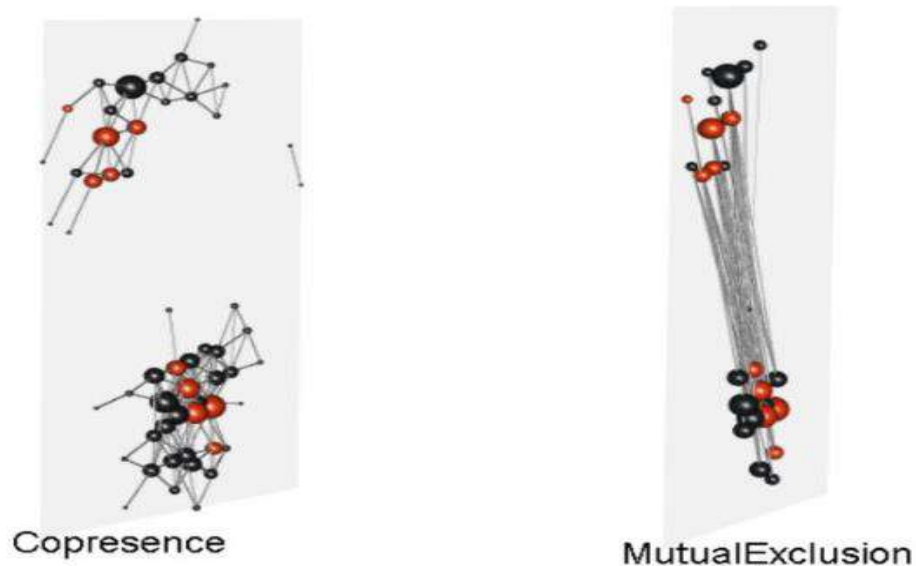


Figura 23: Representación de la microbiota del rumen de bovinos donde se aislaron, a partir de una red original, dos comunidades específicas asociadas a la emisión de metano. Los nodos en rojo representan OTUs cuyas abundancias muestran diferencias significativas entre los individuos con alta y baja emisión de metano. Se

construyó previamente una red de co-presencia, representando las interacciones mutualistas, y una de mutua exclusión, representando las interacciones antagónicas. Solo el 46% de los nodos era compartido entre ambas capas. Los géneros *Roseburia*, *Pseudoramibacter*, *Eubacterium* y *Megasphaera*, mostraron abundancias más bajas en individuos de alta emisión que en el caso de los de baja emisión; además de que sus correspondientes nodos estaban altamente interconectados, lo que sugiere que cumplen un papel importante en la regulación de la emisión de metano. Imagen extraída de Zheng et al., 2018 [21].

3. Objetivo general

Mediante el uso de redes multicapa, el presente trabajo pretende analizar redes de co-abundancia de OTUs bacterianos procedentes de muestras de la microbiota, empleando este enfoque para comparar la arquitectura de las distintas capas y establecer si existen patrones a través de los cuales se pueda obtener información sobre las características de la microbiota, su dinámica y su relación con el o los huéspedes.

3.1 Objetivos particulares

- 1.- Estudiar si existen diferencias en la arquitectura de las redes entre las distintas capas.
- 2.- Identificar comunidades bacterianas dentro de la microbiota que se conserven a lo largo de las capas y analizar si los miembros de dichas comunidades poseen características comunes.
- 3.- Analizar si existen cambios significativos de las abundancias entre capas y si tales cambios pudiesen estar afectando la arquitectura de la red.

4. Materiales y métodos

Se analizaron tres bases de datos tipo *phyloseq* bajo el enfoque de las redes multicapa. La primera de estas contenía datos de la microbiota del escarabajo *Lema daturaphila* y de la microbiota de la planta de la cual este se alimenta durante su estadio larval, la solanácea *Datura inoxia* [100]. También se analizaron dos bases de datos procedentes del paquete *seqtime* [101], correspondientes a series temporales de la microbiota de dos sujetos; el primero de los cuales (sujeto A) realizó un viaje entre los días 71 y 122 durante los cuales cambió su régimen alimenticio, sufriendo en consecuencia dos episodios de diarrea entre los días 80 a 85 y 104 a 113. Por su parte, el sujeto B experimentó una intoxicación por *Salmonella* entre los días 151 y 159 [102].

Para la realización de este trabajo se utilizó la versión 4.1.0 (2021-05-18) de R. Se emplearon los paquetes *igraph* [103], *phyloseq* [104], *SpiecEasi* [105], *minet* [106], *seqtime* [101], *muxViz* [107], *multinet* [108], *corrplot* [109] y *BiodiversityR* [110]. Algunos de los paquetes están disponibles en CRAN y otros fueron descargados de repositorios de GitHub o del proyecto para análisis de datos genómicos Bioconductor.

Paquete	Uso	Fuente	Descarga
igraph	Construcción, diseño, análisis y manipulación de redes.	CRAN	<code>install.packages("igraph")</code>
phyloseq	Herramientas para la importación, análisis, almacenamiento y la visualización de datos de microbioma.	Bioconductor	<code>if(!requireNamespace("BiocManager", quietly = TRUE)) install.packages("BiocManager"); BiocManager::install("phyloseq")</code>
devtools	Provee las funciones necesarias para el desarrollo de paquetes de R y descarga desde repositorios GitHub.	CRAN	<code>install.packages("devtools")</code>
SpiecEasi	Estimación de la covarianza de datos de abundancia e inferencia de redes ecológicas.	GitHub	<code>library(devtools); install_github("zdk123/SpiecEasi")</code>
seqtime	Provee datos de series temporales de microbiota, y herramientas para simular y analizar la dinámica de comunidades microbianas en el tiempo.	GitHub	<code>library(devtools); install_github("hallucigenia-sparsa/seqtime")</code>

muxViz	Construcción, diseño y análisis de redes multicapa.	GitHub	devtools::install_github("manlius/muxViz")
multinet	Construcción, diseño y análisis de redes multicapa.	CRAN	install.packages("multinet")
corrplot	Análisis de la correlación de matrices.	CRAN	install.packages("corrplot")
BiodiversityR	Análisis de diversidad	CRAN	install.packages("BiodiversityR")

Cuadro 1: Paquetes de R utilizados.

También desarrollé una serie de funciones en R que me permitieron analizar los datos, mismas que fueron almacenadas como un paquete de autoría propia que llamamos *ml_BioNets*. Dicho paquete depende de *igraph* [103], *muxViz* [107] y *phyloseq* [104].

Comando	Uso	Inputs
T_collapse(T_table, O_table, names_level)	Suma las abundancias absolutas de los OTUs pertenecientes a cada uno de los clados del nivel taxonómico seleccionado.	Tabla de taxones, tabla de OTUs y nivel taxonómico.
v_colored(g, T_table, p_tax, g_tax, g_colors)	Colorea los nodos en función del nivel taxonómico seleccionado. El resultado se almacena como un atributo asociado a los nodos al que llamamos <i>color</i> .	Red tipo <i>igraph</i> , tabla de taxones, nivel taxonómico superior, nivel taxonómico en el cual se infirió la red, y un carácter que contenga los colores que se asignados a cada clado.
g_abundance(layer_mat, g)	Colorea los nodos en función de su abundancia relativa. El	Matriz de abundancias de la capa y la red de la capa.

TaxGroup(g, T_table, g_tax, p_tax)	<p>resultado se almacena como un atributo asociado a los nodos al que llamamos <i>rel_ab</i>.</p>	<p>Red tipo igrph, tabla de taxones, nivel taxonómico superior, y nivel taxonómico en el cual se infirió la red.</p>
ctr(g.list, ctr_type)	<p>Indica a qué clado, de la categoría taxonómica seleccionada, pertenece un determinado grupo de nodos. El resultado se almacena como un atributo asociado a los nodos al que llamamos <i>Taxon</i>.</p>	<p>Objeto muxViz y tipo de centralidad seleccionada (<i>degree</i>, <i>betweenness</i> o <i>closeness</i>).</p>
ctr_g(g, ctr_type)	<p>Colorea los nodos en un espectro de colores cálidos en función de su centralidad por grado, intermediación o cercanía. El resultado se almacena como un atributo asociado a los nodos al que llamamos <i>hl</i>.</p>	<p>Red tipo igrph y tipo de centralidad seleccionada (<i>degree</i>, <i>betweenness</i> o <i>closeness</i>).</p>
diff_nodes_graph(T_Collapsed, n, mat_list, g.list, alpha)	<p>Colorea los nodos en un espectro de colores cálidos en función de su centralidad por grado, intermediación o cercanía. El resultado se almacena como un atributo asociado a los nodos al que llamamos <i>hl</i>.</p>	<p>Tabla de abundancias de todas las muestras, número a nodos a considerar, lista de matrices de abundancias de las distintas capas, y el nivel de significancia alfa.</p>
diff_nodes_graph(T_Collapsed, n, mat_list, g.list, alpha)	<p>Identifica diferencias significativas entre la media de las abundancias de nodos réplica en redes bipartitas. Esta función sólo considera los nodos con mayor abundancia. El resultado se almacena como un atributo</p>	

<p><code>corr_clusters(clusters, method, as.corrplot)</code></p>	<p>asociado a los nodos al que llamamos <i>colorA</i>.</p> <p>Esta función permite generar una matriz de correlaciones de la clusterización entre las capas de la red, además de un gráfico de correlaciones.</p> <p>Lista de objetos tipo <i>communities</i> para cada capa, generados con alguna de las funciones de clusterización de <i>igraph</i>. Los métodos de comparación de la clusterización entre capas pueden ser “<i>vi</i>”, “<i>nmi</i>”, “<i>split.join</i>”. “<i>rand</i>” o “<i>adjusted.rand</i>”. Si <i>as.corrplot = T</i>, la función genera un gráfico de correlaciones, de lo contrario arroja una matriz de correlaciones.</p>
--	--

Cuadro 2: Funciones del paquete *ml_BioNets*.

Previo a la inferencia de las redes, se colapsaron los datos a nivel de género en todas las bases de datos, eliminando previamente todos aquellos no-clasificados. Para esto se construyó la función *T_Collapse()*, que emplea como entrada la tabla de OTUs, la tabla de taxones y el nivel taxonómico hasta el cual se desea colapsar los datos. Dicha función suma las abundancias absolutas de los OTUs pertenecientes a cada uno de los clados dentro de la categoría taxonómica seleccionada.

Para inferir las redes se emplearon como datos de entrada una tabla de OTUs, que contiene las abundancias relativas de los OTUs en las muestras, y una tabla de taxones, misma que contiene la clasificación taxonómica de los OTUs. Para generar las redes multicapa previamente se separaron los datos concernientes a cada una de las capas. Para el caso de la red planta-insecto se separaron los datos correspondientes al escarabajo (huevos, intestino y excremento) y aquellos pertenecientes a la solanácea (endófitos, epifitos y semillas). Para los datos de series temporales, se separaron los períodos de salud y de enfermedad. Se aislaron tres secciones distintas de los datos del sujeto A, tomando en cuenta el periodo previo al viaje, durante el viaje y después del viaje. No fue posible construir redes únicamente a partir de los datos de los dos periodos de diarrea durante el viaje, debido a que el número de muestra era demasiado bajo como para que cualquiera de los algoritmos posteriormente utilizados pudiese inferir correlaciones entre los nodos. Para el sujeto B se tomó en cuenta el periodo previo a la infección por *Salmonella*, el periodo durante el cual se desarrolló la infección, y el periodo posterior a esta.

Posteriormente se procedió a la inferencia de las redes para cada una de las capas, para lo cual se emplearon los algoritmos SparCC del paquete SpiecEasi [105] y ARACNe del paquete *minet* [106]. Para el caso del algoritmo SparCC, se eliminaron todas las correlaciones por debajo de un umbral de 0.4. Inicialmente se planteó utilizar también el algoritmo SPIEC-EASI, pero debido a su alta selectividad, no fue posible construir redes para algunas de las capas con bajos números de muestra, razón por lo cual su uso se descartó. Se analizó también la diversidad de las distintas capas mediante los índices de Shannon y Simpson, además de la abundancia relativa de los *phyla* en las capas. Se empleó el índice de diversidad de Rényi, igualmente conocida como entropía de Rényi, mismo que cuantifica la incertidumbre del sistema en base al número de taxones y sus abundancias absolutas [111].

Para visualizar la distribución de los *phyla* en la red se generó la función `v_colored()`, que asigna un color a todos los nodos que pertenezcan a un mismo filo. Se construyó la función `g_abundance()` con el fin de visualizar los nodos de la red en función de su abundancia relativa. Para analizar las características de la red también se analizó la distribución del *degree* de las capas con una prueba de Shapiro-Wilk, y se comparó la distribución del *degree* entre las capas.

El análisis de *clusters* se realizó por el método *Louvain*, mismo que optimiza la modularidad para detectar comunidades de nodos muy interconectados mediante el cálculo la densidad de aristas dentro de la red, cuyos valores pueden variar entre 1 y -1 [112]. También se emplearon como complemento los métodos *optimal* y *fast-greedy*, para analizar la conservación de las comunidades encontradas bajo distintos algoritmos [113]. El primero maximiza la modularidad de todas las particiones posibles para calcular la estructura óptima de las comunidades del grafo [114], mientras que el segundo busca regiones con alta densidad de nodos y aristas en el grafo [115].

Se analizó la similaridad de los *clusters* de las capas mediante el índice Rand, cuyo valor varía entre 0 y 1, y que toma en cuenta todos los pares de nodos posibles presentes y no presentes en el mismo *cluster* en dos comunidades distintas, como ratio del número de pares desordenados posibles [116]. También se utilizó el coeficiente NMI (*Normalized Mutual Information*), cuyo valor también oscila entre 0 y 1, que calcula la información compartida entre *clusters* en función de los elementos compartidos, y considerando además la entropía como el número de clases de elementos distintos contenidos en un *cluster* [117].

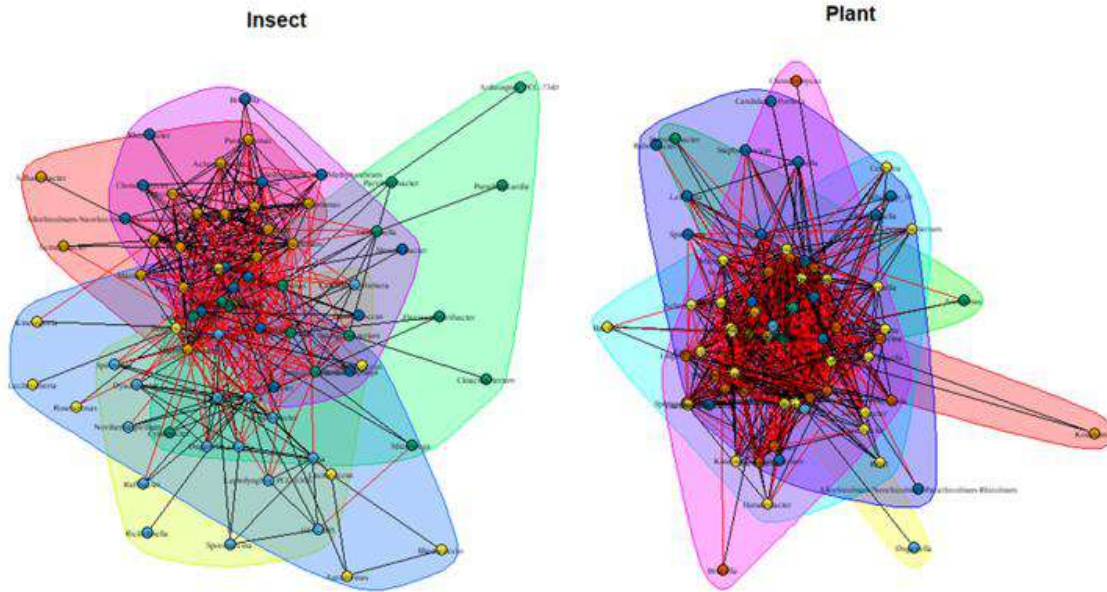


Figura 24: Clusterización mediante el método Louvain de las capas de la red escarabajo-solanácea. Los *clusters* corresponden a regiones de la red ampliamente interconectadas. Las aristas que conectan los nodos dentro de la comunidad están representadas en negro, mientras que aquellas que conectan nodos entre comunidades distintas resaltan en rojo. Para facilitar la visualización de las comunidades se eliminaron los nodos cuyo *degree* fuese nulo.

Se construyeron las redes multicapa utilizando el paquete *muxViz*. La red escarabajo-solanácea se analizó bajo el enfoque de las redes multiplex, mientras que las redes de los sujetos A y B se analizaron como redes temporales. Se compararon los nodos de las comunidades entre las capas, para lo cual se utilizó la función `TaxGroup()`, que nos permitió identificar el *phylum*, clase y familia a la que pertenecen los conjuntos de nodos clusterizados.

Se analizó la relevancia de los nodos de las comunidades mediante distintas medidas de centralidad, para lo cual se construyeron las funciones `ctr()` y `ctr_g()` para redes multicapa y las redes simples, respectivamente. Solo se consideraron las centralidades por *degree* e intermediación, debido a que la centralidad por cercanía no se definió correctamente ante la existencia de nodos y comunidades aisladas. Se identificaron nodos réplica con y sin diferencias significativas entre pares de capas mediante una prueba de t-Student ($p < 0.05$). Únicamente se consideraron los 20 nodos con mayor abundancia dentro de las muestras. Finalmente, mediante el paquete *multinet* [108] se analizó la correlación entre capas con respecto a los nodos, el *degree* y las conexiones en las redes temporales de los sujetos A y B.

5. Resultados

5.1 Resultados de la red escarabajo-solanácea

De los tres algoritmos empleados en la inferencia de redes, solo ARACNe y SparCC fueron empleados en la construcción de las redes multicapa, descartando el uso de SPIEC-EASI para posteriores análisis. Este último está basado en la independencia condicional de los nodos, por lo cual es mucho más selectivo, y dado el bajo número de muestra de algunas de las capas, el algoritmo arroja redes inconexas para tales casos. El análisis de la distribución del *degree* de las capas reveló que esta se asemeja a una distribución normal. Cabe destacar que la interconectividad de la red del escarabajo realizada con ARACNe presentó una interconectividad considerablemente menor a la de su homóloga realizada con SparCC.

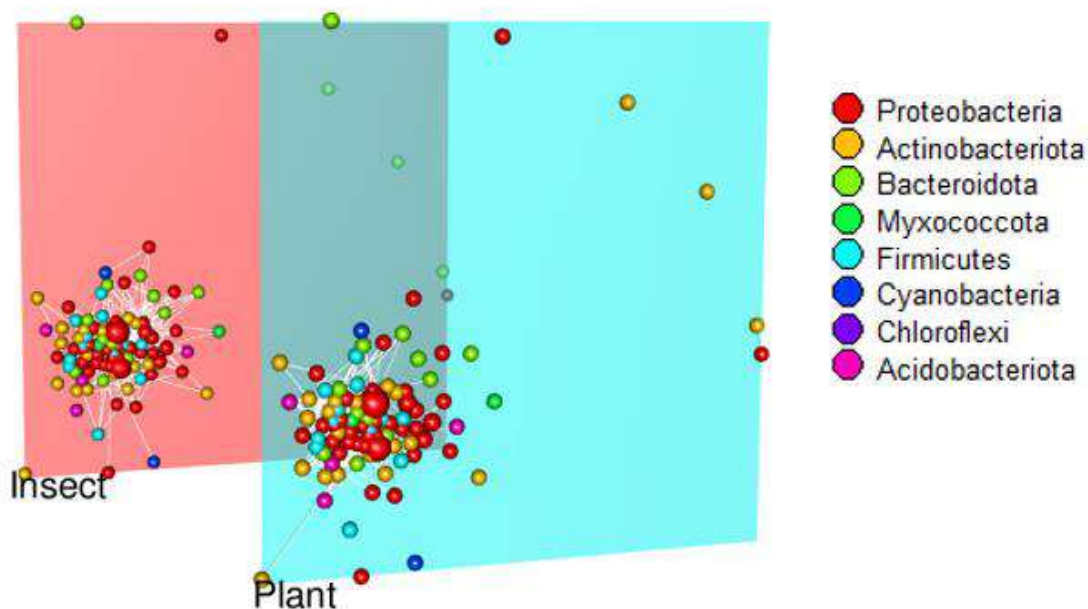


Figura 25: Red multiplex (SparCC) de escarabajo-solanácea. El tamaño de los nodos está en función de la abundancia relativa de los nodos, cuyo color está asociado al *phylum* al que pertenecen.

Las redes multicapa construidas con SparCC y ARACNe de escarabajo-solanácea, en ambos casos, revelaron a Proteobacteria como el *phylum* con mayor abundancia y distribución en las capas. Del mismo modo, su abundancia relativa en las muestras de tejidos fue preponderante.

En las muestras extraídas del escarabajo se encontró una gran abundancia de las gamma-proteobacterias *Serratia* y *Pseudomonas*, pero una abundancia muy baja para el resto de OTUs presentes. La microbiota de la solanácea mostró una mayor diversidad de OTUs con abundancias altas, dentro de los que podemos destacar gamma-proteobacterias como *Pseudomonas*, *Pantoea*, *Siccibacter*, *Janthinobacterium* y *Massilia*.

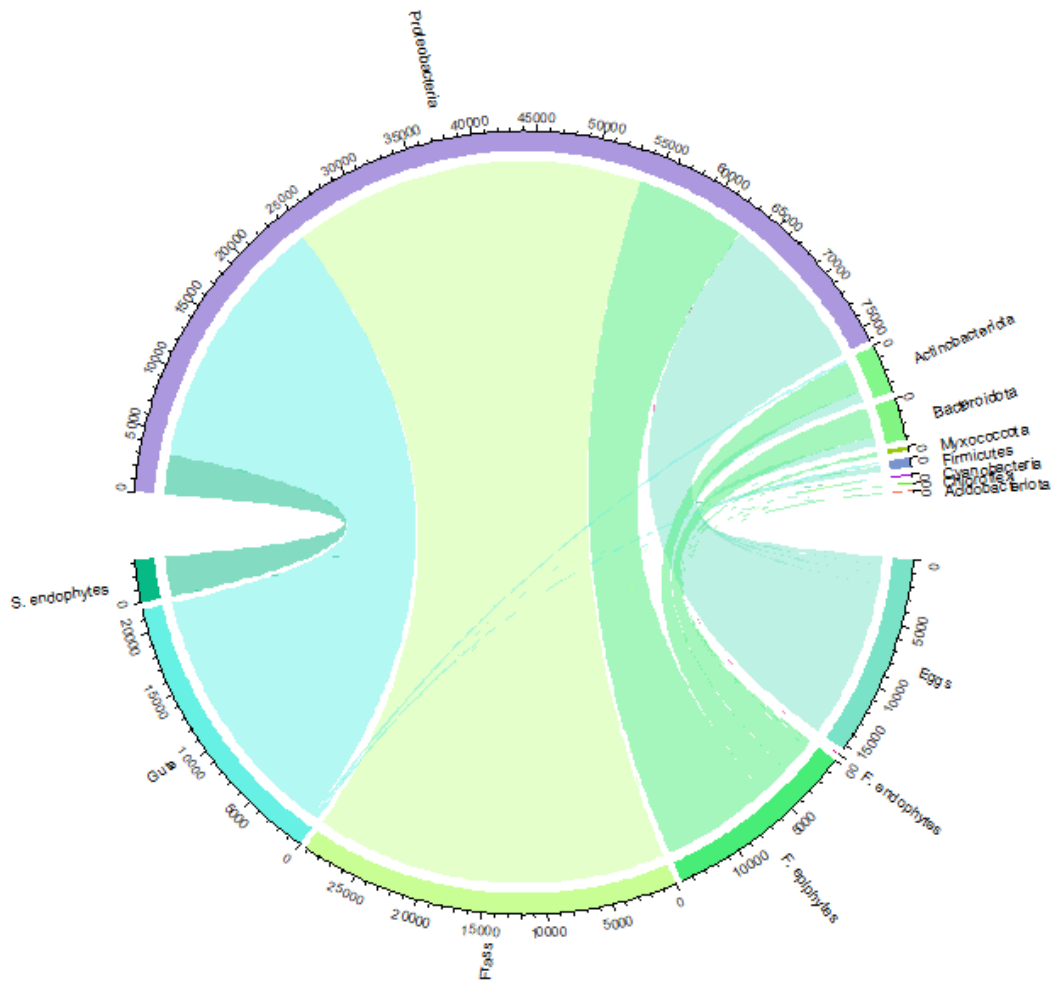


Figura 26: Se analizó la distribución de los *phylum* en las muestras de los distintos tejidos, mostrando una presencia dominante de Proteobacteria en todas las muestras, seguido en menor medida por Actinobacteria y Bacteroidetes.

Las muestras de excremento e intestino de *Lema daturaphila* mostraron una diversidad de *phyla* muy baja ante la abundancia dominante de Proteobacteria. Lo mismo sucede con las muestras endófitas extraídas de la planta. Las muestras con mayor diversidad fueron aquellas extraídas de los huevos del insecto y las muestras de epífitas de la solanácea, que también mostraron una presencia importante de Actinobacteria y Bacteroidetes, y en menor medida también de Firmicutes. Estas muestras, incluyendo las de excremento, obtuvieron valores superiores a 0.4 con el índice de Shannon, lo que indica una complejidad biológica superior a la del resto de muestras. Sin embargo, dichas muestras junto a las de intestino, mostraron valores cercanos a 1 en la diversidad de Simpson. Lo anterior significa que, si bien algunas muestras mostraron abundancias significativas de distintos taxones, la dominancia de las proteobacterias sigue siendo preponderante en las muestras.

El análisis de centralidades por *degree* reveló que, a pesar de la variación de las abundancias entre ambos organismos, las Proteobacterias, conservan un alto grado de centralidad entre capas, lo que sugiere que podrían tener una importancia a nivel funcional en la relación entre el escarabajo y su fuente de alimento. Dentro de este grupo podemos destacar a *Siccibacter* y *Pantoea* en la microbiota del escarabajo, mismas que también mostraron previamente una abundancia relativa alta. El género *Salmonella* también mostró un *degree* alto en la red SparCC, aunque no se encontrara dentro de los OTUs con mayor abundancia relativa.

En el caso de la solanácea, la diversidad de especies con un *degree* elevado fue mucho mayor. Destacan alfa-proteobacterias como *Sphingomonas*, *Skermanella*, *Methylobacterium* y *Microvirga*, además de gamma-proteobacterias como *Janthinobacterium* y *Massilia*. En menor proporción se encontraron también bacterias ajenas al *phylum* Proteobacteria como *Actinoplanes*, *Hymenobacter* y *Cystobacter*, pertenecientes a Actinobacteria, Bacteroidetes y Myxococcota, respectivamente.

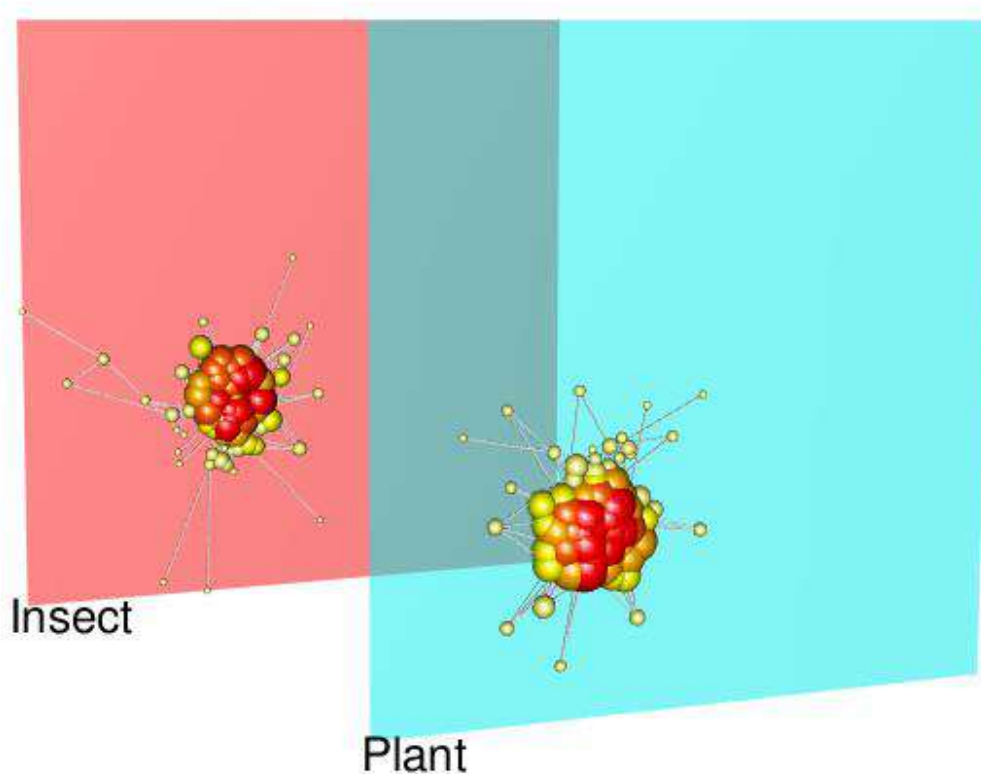


Figura 27: Red multiplex de la centralidad por *degree* entre las capas. Nodos con tonos más próximos al rojo indican un mayor número de conexiones.

El análisis de centralidad por cercanía mostró valores altos para casi todos los nodos debido a la alta interconectividad de los *clusters*. Del mismo modo, no se encontró consenso al respecto de la centralidad por *betweenness* entre los dos algoritmos.

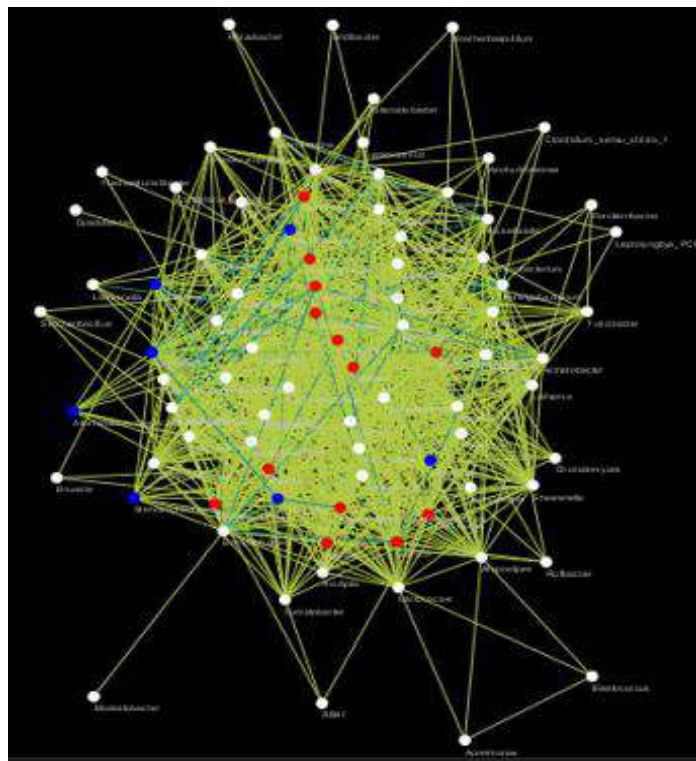


Figura 28: Gráfico de la variación de los nodos entre las capas de la red bipartita. Los nodos marcados en rojo son aquellos que poseen diferencias significativas en las abundancias entre capas, mientras que los azules son aquellos que no poseen diferencias significativas entre ambas redes. Las conexiones representan la suma de interacciones en ambas capas.

No se encontraron diferencias significativas ($p < 0.05$) en los géneros *Kosakonia*, *Stenotrophomonas*, *Pseudomonas*, *Azomonas*, *Rhizobium*, *Achromobacter* y *Enterobacter*, todos pertenecientes al phylum Proteobacteria. Por el contrario, las proteobacterias *Serratia*, *Siccibacter*, *Escherichia*, *Shigella*, *Pantoea*, *Janthinobacterium*, *Methylobacterium* y *Salmonella*, además de *Bacillus* de Firmicutes, se encontraron con mayor abundancia en los tejidos del escarabajo. Sin embargo, las proteobacterias *Sphingomonas*, *Massilia*, *Rubellimicrobium* y *Skermanella*, junto a *Hymenobacter* de Bacteroidetes, tenían una abundancia mayor en los tejidos vegetales.

5.2 Resultados de las redes para los sujetos A y B

Al igual que en el caso anterior, las redes fueron generadas únicamente con los algoritmos ARACNe y SparCC, pues no fue posible inferir las redes con SPIEC-EASI para la segunda capa de ambos sujetos debido a que, al tener un número de muestra bajo, el algoritmo no fue capaz de encontrar correlaciones entre sus componentes. La interconectividad de las redes generadas con ARACNe fue menor que en las redes generadas con SparCC. El análisis de la distribución del *degree* en las redes ARACNe reveló que se trata de redes de mundo pequeño, donde la mayoría de las conexiones están concentradas en unos pocos nodos y la mayoría de los nodos tiene pocas conexiones. Ninguna de las redes SparCC presentó una distribución normal. En el caso de la diversidad Simpson como en la diversidad de Shannon, solo las muestras

correspondientes a los periodos de enfermedad poseían una distribución normal en contraste con los estadios basales, lo que podría indicar un cambio en la dinámica de las poblaciones al presentarse la disbiosis.

Tanto en el sujeto A como en el sujeto B, la presencia de Firmicutes y Bacteroidetes era preponderante sobre el resto de *phyla*, razón por la cual todas las muestras obtuvieron valores cercanos a la unidad en la prueba de diversidad de Simpson, en contraste con los valores bajos de diversidad de Shannon. En el sujeto A hubo un aumento significativo en la abundancia relativa de las proteobacterias en detrimento de Firmicutes, así como una reducción de Actinobacteria, durante el periodo de viaje. Durante el periodo de infección por *Salmonella* en el sujeto B, también se registró un descenso en la abundancia relativa de Firmicutes, en contraste con un aumento en Bacteroidetes. Al igual que en el sujeto A, se identificó una reducción en la proporción de Actinobacteria. A pesar de que tras el periodo de infección hay un aumento de Firmicutes con respecto a Bacteroidetes, no se recuperan los valores previos y la abundancia relativa de Actinobacteria continúa siendo baja con respecto a sus valores iniciales.

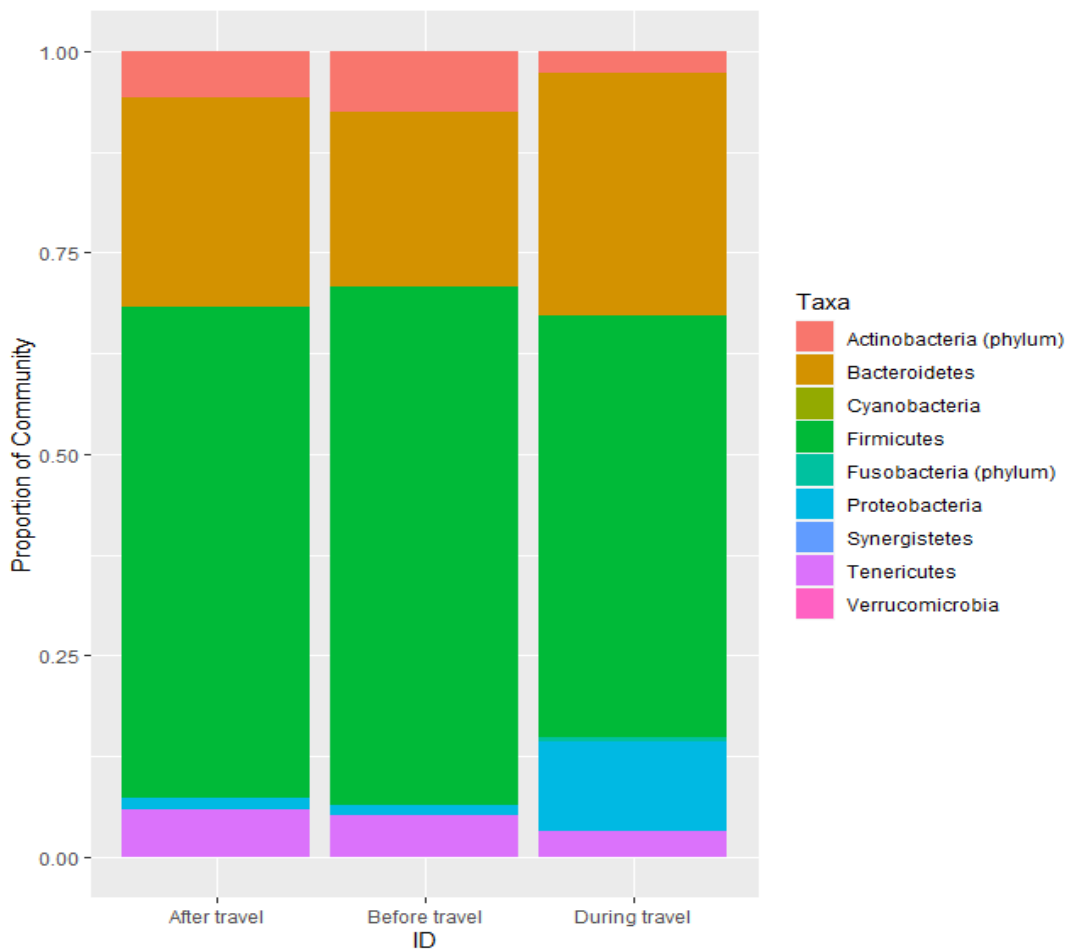


Figura 29: Distribución de las abundancias de los *phyla* en el sujeto A en las tres capas, correspondientes al periodo previo al viaje, durante el viaje y posterior al viaje. Es durante el periodo de viaje en el cual el sujeto sufrió una alteración en la composición de su microbiota, derivada de un cambio en su dieta habitual.

En el sujeto A los géneros *Bacteroides* y *Faecalibacterium* se mantuvieron como los más abundantes en todas las capas. También se detectaron abundancias altas de *Clostridium* y *Blautia*. La abundancia de *Bifidobacterium* decayó durante el periodo de disbiosis, aunque se recuperó después, a diferencia de las abundancias de *Coprococcus*, cuya población no recuperó las abundancias iniciales. Durante la disbiosis se detectó un aumento considerable en las abundancias de *Escherichia*, mismas que se redujeron tras el viaje. Por su parte, en el sujeto B, a pesar de tratarse de los dos géneros con mayor abundancia relativa, la presencia de *Bacteroides* en las muestras fue mucho mayor que la de *Faecalibacterium*. Si bien, las abundancias de ambos géneros se mantuvieron constantes, hubo un aumento en la abundancia relativa de *Faecalibacterium* en el periodo post-infección. En las redes SparCC los géneros *Bacteroides*, *Faecalibacterium*, *Clostridium* y *Blautia* mantuvieron valores altos de centralidad por *degree* en todas las capas de ambas redes temporales.

Las redes generadas con ARACNe presentaron comunidades con nodos altamente interconectados pero aislados del resto de nodos de la red, una particularidad que solo se presentó en las capas correspondientes a la disbiosis, en contraste con una mayor interconectividad de toda la red en las demás capas. La presencia de dichas comunidades se conservó tras analizar el agrupamiento de las capas con los algoritmos *Louvain*, *optimal* y *fast-greedy*. Cabe destacar que en el sujeto B, tras el periodo de disbiosis, la arquitectura de la red no se recuperó al encontrarse los nodos menos interconectados que al inicio del muestreo, en contraste con lo sucedido en el sujeto A.

En las redes ARACNe del sujeto A destacaban dos grandes comunidades altamente interconectadas en la capa de disbiosis. Ambas comunidades tenían una importante presencia de proteobacterias como *Thauera*, *Pseudoxanthomonas*, *Brachymonas*, *Polynucleobacter*, *Rhodoplanes*, *Massilia*, *Methylotenera* y *Rheinheimera*. Cabe destacar que las beta-proteobacterias fueron la clase más abundante en este grupo. Bacteroidetes también tuvo una presencia importante en dichas comunidades, destacando géneros como *Hymenobacter*, *Flavisolibacter*, *Cytophaga* y *Pedobacter*. Firmicutes también tuvo presencia con géneros como *Alicyclobacillus*, *Paenibacillus* y *Caloramator*. Finalmente podemos mencionar la presencia de algunas actinobacterias como *Rathayibacter*, *Microcella* y *Yonghaparkia*.

En los estadios basales del sujeto A, solo se encontraron comunidades pequeñas de entre tres y seis elementos. De la primera capa podemos destacar una comunidad formada por los géneros *Rhizobium*, *Erysipelothrix*, *Williamsia*, *Marinilactibacillus*, *Cronobacter* y *Providencia*. En la tercera capa no se encontraron comunidades mayores a cuatro elementos. Tales comunidades no se conservaron en las redes SparCC, donde la mayoría de los OTUs identificados como pertenecientes a las comunidades encontradas en las redes ARACNe, se hallaron como nodos aislados en las redes SparCC. La única excepción a esta regla fueron las comunidades de la segunda capa, cuyos elementos estaban integrados a comunidades de mayor tamaño en las redes SparCC.

En análisis de clusterización de las redes SparCC del sujeto A mostró grandes comunidades de bacterias, cuya composición se mantuvo similar a lo largo de todas las capas a pesar de los cambios en las abundancias relativas de las muestras. Géneros muy

abundantes en las muestras como *Bacteroides*, *Faecalibacterium*, *Clostridium* y *Blautia* obtuvieron valores altos de *degree*, lo que podría indicar que poseen una importancia destacable dentro de las comunidades encontradas. Cabe destacar que si bien, las comunidades encontradas tenían una presencia importante de Firmicutes y Proteobacteria, formando ambos grupos dos comunidades de gran tamaño, la presencia de Bacteroidetes, a pesar de su abundancia relativa alta en contraste con la mayoría de *phyla*, fue marginal. Es importante destacar que a pesar de la reducción en las abundancias de Firmicutes y el aumento de Proteobacteria y Bacteroidetes durante el periodo de disbiosis, Bacteroidetes continuó teniendo una presencia marginal en las clusterizaciones encontradas.

Sin embargo, los algoritmos NMI y Rand empleados para analizar la correlación de la clusterización entre capas no arrojaron diferencias significativas, lo que podría estar asociado al gran número de nodos independientes presentes tanto en las redes SparCC como en las de ARACNe.

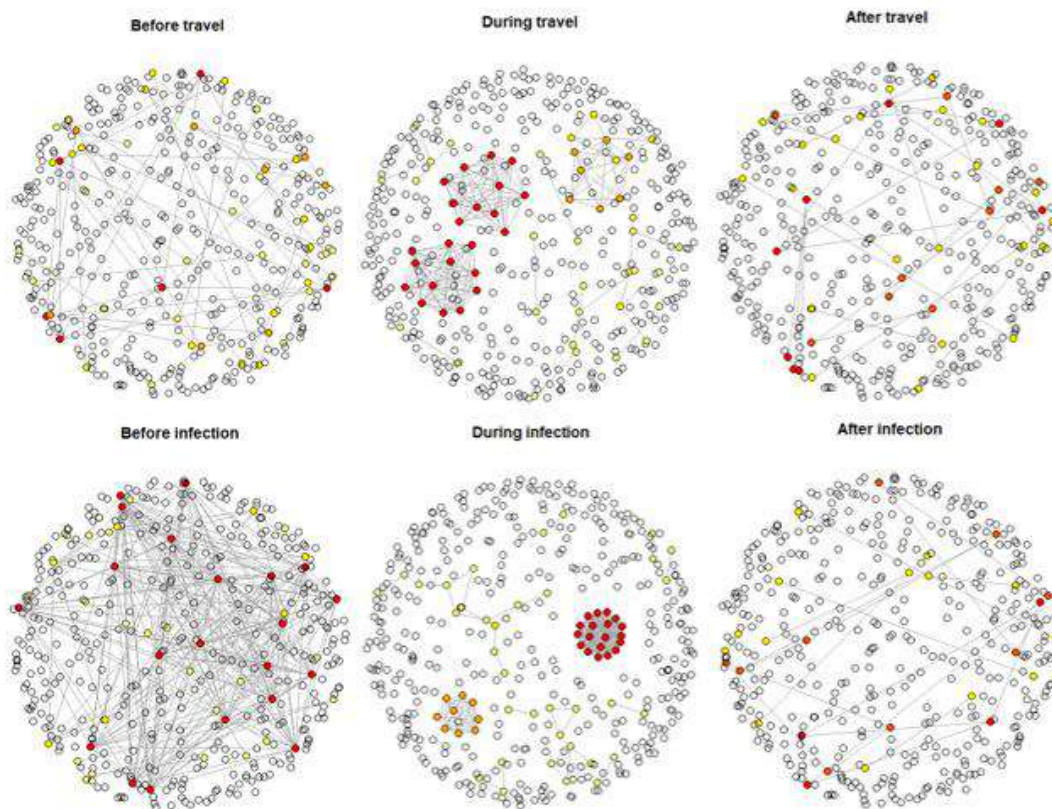


Figura 30: Capas de las redes temporales de los sujetos A y B. Las regiones de la segunda capa de las redes ARACNe marcadas en rojo, corresponden a zonas con una alta densidad del *degree*, donde fueron halladas comunidades de bacterias altamente interconectadas compuestas principalmente por géneros de Firmicutes y Proteobacteria.

En el sujeto B, se identificó una reducción significativa en las abundancias de *Bacteroides*, *Actinobacillus*, *Phascolarctobacterium*, *Clostridium*, *Streptococcus*, *Oscillospira*, *Alistipes* y *Parabacteroides* en el sujeto A entre la primera y la segunda capa,

además de un aumento de *Erwinia* durante el periodo de disbiosis. Las abundancias de dichos géneros recuperan sus proporciones originales tras el periodo de disbiosis. En el sujeto B el cambio en las abundancias fue mayor, especialmente entre la segunda y la tercera capa. De los géneros que se vieron modificados, podemos destacar a *Bacteroides*, *Hylemonella*, *Actinobacillus*, *Erwinia*, *Phascolarctobacterium*, *Clostridium*, *Ruminococcus*, *Coprococcus*, *Eubacterium*, *Faecalibacterium* y *Akkermansia*. A diferencia de lo sucedido con el sujeto A, en este caso las abundancias de los géneros no solo no se recuperaron tras el periodo de disbiosis, sino que la mayoría de estos presentaron una abundancia incluso menor a la registrada durante el periodo de infección.

A pesar de la reducción de Firmicutes durante la infección por *Salmonella*, la presencia de dicho *phylum* fue destacable en los *clusters* hallados en las redes ARACNe de la capa de disbiosis. Dentro de los géneros de Firmicutes identificados como pertenecientes a las comunidades del periodo de disbiosis, se incluyen *Peptococcus*, *Peptostreptococcus*, *Facklamia*, *Lachnospira*, *Vagococcus*, *Clostridium*, *Blautia*, *Ruminococcus*, *Anaerostipes*, *Eubacterium*, *Weissella* y *Faecalibacterium*. También destacó la presencia de proteobacterias como *Vogesella*, *Vitreoscilla*, *Serratia*, *Desulfovibrio*, *Aquamonas*, *Methyloversatilis*, *Aquitalea* y *Microvirgula*. La presencia de Bacteroidetes y Actinobacteria en tales comunidades, como también se observó en el sujeto A, fue marginal a pesar de la alta abundancia relativa de ambos *phyla* en las muestras.

En la primera capa del sujeto B se encontró una macrocomunidad compuesta por un gran número de géneros, en contraste con el sujeto A donde solo había comunidades pequeñas. Al igual que con las comunidades encontradas en las redes del sujeto A, la presencia de Proteobacteria y Firmicutes fue preponderante en todos los *clusters*, un patrón presente tanto en las redes ARACNe como en las redes SparCC. Dentro de las proteobacterias halladas en los *clusters* de la primera capa de las redes generadas por ambos algoritmos, destacamos los casos de *Vogesella*, *Vitreoscilla*, *Serratia*, *Desulfovibrio*, *Methyloversatilis* y *Aquamonas*. De los géneros pertenecientes a Firmicutes podemos destacar a *Arcanobacterium*, *Peptococcus*, *Peptostreptococcus*, *Facklamia*, *Lachnospira* y *Vagococcus*, mismos que también estuvieron presentes en los *clusters* de la red SparCC. En ambos casos, en la tercera capa se encontraron comunidades de menor tamaño que las halladas en la primera capa, sin embargo, a diferencia del caso anterior no se encontró un consenso en cuanto a los elementos constituyentes de los *clusters* hallados con ambos algoritmos. A pesar de ello, cabe destacar que tales comunidades contenían una presencia preponderante de Firmicutes y Proteobacteria. Al igual que en las redes del sujeto A, la presencia de Bacteroidetes en las comunidades del sujeto B fue marginal. Sin embargo, sí se encontró una presencia importante en todas las capas de Actinobacteria, destacando géneros como *Arcanobacterium*, *Rothia*, *Microvirgula*, *Bifidobacterium* y *Varibaculum*, presentes en las comunidades halladas con ambos algoritmos. Tanto en las redes ARACNe como en las SparCC, en la segunda capa, referente al periodo de infección, se encontraron comunidades de mayor tamaño, comparadas con las capas del estado basal. Tales comunidades, sin embargo, presentaron una presencia importante de Firmicutes y Proteobacteria, como en los casos anteriores, además de una presencia considerablemente menor de Bacteroidetes.

Las alfa-proteobacterias mencionadas se hallaron agrupadas en un *cluster* junto a otras bacterias como *Hymenobacter*, *Pseudomonas*, *Janthinobacterium* y *Massilia*. Además, todas las mencionadas se encontraron solo en las muestras de epífitas, lo que sugiere que forman una comunidad bacteriana cuyos miembros podrían tener funciones biológicas comunes. Sin embargo, la escasa diversidad de bacterias detectadas en las muestras de semillas y endófitas, también sugiere que el hecho de que la superficie de la planta podría simplemente ser un sitio más propicio para la proliferación de bacterias que los otros dos tejidos. Debido a ello, se requiere de un análisis mucho más exhaustivo que involucre agrupar los nodos por su función biológica para verificar si los *clusters* realmente corresponden a comunidades relacionadas por su actividad biológica.

6. Discusión

6.1 Discusión de la red escarabajo-solanácea

La relación de *Lema daturaphila* y las especies del género *Datura* es ampliamente conocida, puesto que dicho escarabajo es considerado como una plaga para múltiples especies de dicho género. A pesar de que *Datura innoxia* sea una planta tóxica para la mayoría de los insectos herbívoros, este escarabajo se ha adaptado para alimentarse de dicha planta solanácea. Las larvas del escarabajo se alimentan de las hojas de la planta durante el verano, y a menudo se les puede encontrar concentradas en grandes grupos en la superficie de las hojas [118]. Sin embargo, nuestro conocimiento sobre el papel de la microbiota en la relación entre ambas especies y su interacción con sus respectivos huéspedes, es aún muy limitado.

Todos los análisis revelaron una amplia presencia del *phylum* Proteobacteria, un clado que sin embargo es muy diverso y, por lo tanto, difícil de asociar a una única función concreta. Su relación con la planta y el insecto varía a lo largo de un espectro muy amplio que incluye especies comensales, patógenas, simbioses y bacterias de vida libre.

Dentro de este marco cabría señalar a las alfa-proteobacterias, algunas de las cuales suelen estar asociadas a tejidos vegetales como bacterias fijadoras de nitrógeno, un papel que también desempeñan varias especies de beta-proteobacterias, clase que, sin embargo, no se encontró en proporciones significativas. Algunas de las alfa-proteobacterias identificadas, tales como *Skermanella* y *Microvirga*, se encuentran en una gran variedad de hábitats, especialmente en el suelo [119], lo que en primera instancia nos llevaría a plantear la hipótesis de que se trata de bacterias de vida libre. Sin embargo, el hecho de que estas se concentran casi exclusivamente en las muestras de epífitas sugiere una relación simbiótica o de comensalismo.

En el caso de *Sphingomonas* y *Methylobacterium*, ambos son géneros de bacterias fijadoras de nitrógeno, el primero de los cuales suele ser más abundante en las raíces de las plantas, mientras que en el género *Methylobacterium* podemos encontrar

simbiontes endófitas como *Methylobacterium symbioticum*, y simbiontes epífitas como *Methylobacterium extorquens* [120], aunque en las muestras analizadas solo se encontraron en las muestras de epífitas. Esto sugiere que la alta abundancia relativa de alfa-proteobacterias en las muestras, especialmente de *Skermanella* debido a su sobresaliente presencia, podría estar relacionado con una asociación de tipo mutualista o comensal, como bacterias fijadoras de nitrógeno, con los tejidos superficiales de la solanácea.

Lo anterior contrasta con la mayor diversidad encontrada en las muestras del escarabajo. En dicho caso, la presencia de bacterias particularmente abundantes en heces e intestinos, como *Serratia*, *Pseudomonas*, *Siccibacter*, *Pantoea* y *Salmonella*, también en muestras huevos, podría explicarse debido a que las larvas suelen cubrirse de excremento como mecanismo protector contra los depredadores [121]. Este fenómeno también podría explicar por qué se encontraron estas mismas bacterias en la superficie de la planta a pesar de estar fuertemente asociadas al intestino del escarabajo. Algunas de estas bacterias podrían estar involucradas en la digestión, pero también en la protección contra patógenos, como sucede con *Serratia symbiotica* y su asociación con los áfidos, lo que le otorga a este último protección contra parásitos [122]. Esta hipótesis se ve fomentada ante el hecho de que *Serratia* junto a *Pseudomonas* son los géneros más abundantes en las muestras, y están ampliamente distribuidas en todos los tejidos analizados del insecto.

El análisis mediante redes multicapa identificó conjuntos de bacterias que podrían, dados los datos arrojados por los análisis, tener una función biológica importante en la regulación de la propia red y algunos incluso podrían actuar como simbiontes de sus huéspedes. Sin embargo, se requiere compaginar esta información con una base de datos de modo que podamos categorizar y etiquetar los nodos, e identificar si las funciones comunes asociadas a los OTUs se corresponden con los *clusters* hallados.

6.2 Discusión de las redes temporales de los sujetos A y B

La microbiota intestinal humana es sumamente estable. A pesar de la constante variación en las abundancias de las poblaciones bacterianas a lo largo del tiempo, la mayor parte de estas poblaciones tiende a mantenerse dentro de un rango, independientemente de las constantes perturbaciones derivadas de la ingesta de alimentos, medicamentos, actividad física, entre otros factores involucrados [123]. Enfatizando lo anterior, en el estudio de Caporaso *et al.* (2011) se calculó que alrededor del 95% de los OTUs bacterianos encontrados en muestras de intestino y saliva, permanecieron en rangos estables a lo largo del periodo de muestreo, correspondiente a un año [124]. En general, a pesar de las constantes fluctuaciones en las poblaciones bacterianas a lo largo de una serie de tiempo, estas tienden a aproximarse a determinados valores de equilibrio que dependen en gran medida de sus interacciones con otros OTUs bacterianos, lo que mantiene la estabilidad de las comunidades. Sin embargo, la exposición prolongada y sostenida a un agente perturbador, como puede ser un antibiótico, si promueve cambios importantes en la dinámica de las poblaciones a largo plazo [125].

El estudio realizado por David *et al.* (2014) mediante el enfoque de las series temporales, reveló que, en el periodo de viaje durante el cual el sujeto A vio modificado su régimen alimenticio, la dinámica de las poblaciones bacterianas se vio alterada significativamente. Sin embargo, tras dicho periodo, y luego de que el sujeto A retomara su dieta habitual, la dinámica de poblaciones se restableció a un estado similar al registrado previamente al viaje. En contraste, en el sujeto B un fenómeno similar se observó durante el periodo de infección por *Salmonella*, sin embargo, a diferencia del primer caso la dinámica de la microbiota no se recuperó, y se reportó una considerable pérdida de la diversidad [126].

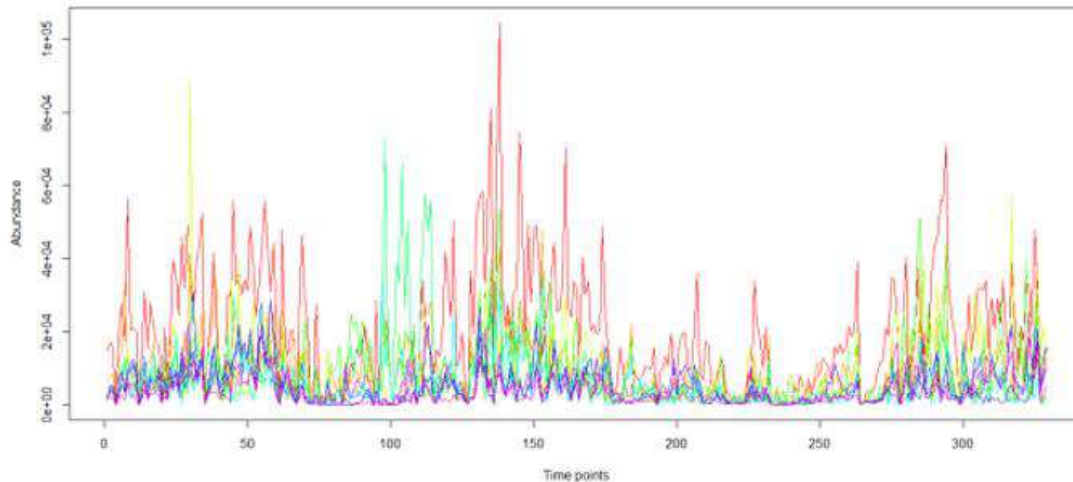


Figura 31: Serie temporal de los diez OTUs más abundantes en las muestras del sujeto A, mismo que sufrió un cambio en la dinámica de su microbiota derivado de un viaje realizado entre los días 71 y 122, durante los cuales modificó su dieta habitual.

El cambio en la distribución tanto del *degree*, en las redes ARACNe, como en la distribución de la diversidad en los periodos de disbiosis en contraste con los estadios basales, podría ser un indicativo de una alteración en la dinámica habitual de las poblaciones bacterianas. Sin embargo, en las redes SparCC del sujeto A, la evidencia en cuanto a un cambio en la interconectividad y los *clusters* encontrados en la red no arrojó resultados concluyentes, a pesar de las alteraciones en las abundancias encontradas en las muestras. Sin embargo, el cambio en la arquitectura de las redes ARACNe, especialmente en lo referente a las comunidades bacterianas halladas en periodos de disbiosis en contraste con los periodos basales, sugiere una posible alteración de las interacciones entre los OTUs bacterianos que potencialmente podrían contribuir a entender el papel de la microbiota en periodos de enfermedad y como esta misma microbiota se ve alterada en tales períodos.

Si bien en el sujeto B, tanto las abundancias como la arquitectura de la red ARACNe no parecen recuperarse tras el periodo de disbiosis, la ausencia de datos posteriores no permite descartar la posibilidad de que la microbiota del sujeto B no se haya recuperado en los días posteriores, o si realmente la infección por *Salmonella* derivó en cambios significativos que perdurasen a largo plazo.

En ambos casos se encontró una alteración en el ratio Bacteroidetes/Firmicutes (B/F), reflejándose una reducción de la abundancia relativa de Firmicutes en detrimento de Bacteroidetes y de otros *phyla*, como es el caso de Proteobacteria en el sujeto A. La importancia de los *phyla* Firmicutes y Bacteroidetes en la microbiota humana es bien conocida, pues constituyen alrededor del 90% de la diversidad microbiana intestinal [127]. Esto ayudaría a explicar los valores altos en el índice de Simpson de todas las muestras, pues se trata de un índice de dominancia cuyos valores se incrementan en presencia de taxones con abundancias relativas predominantes en comparación con el resto de las poblaciones [128]. Si bien, la abundancia de estos dos *phyla* puede variar entre individuos dependiendo de sus características fisiológicas, su estilo de vida, alimentación o factores genéticos, se considera que un aumento en la abundancia relativa de Bacteroidetes en detrimento de las poblaciones de Firmicutes está asociado al síndrome de intestino irritable. El caso contrario se ha relacionado con la obesidad [127]. En el caso de los sujetos A y B, la alteración del ratio B/F podría estar asociada al proceso inflamatorio derivado de un proceso infeccioso [129]. Si bien en el sujeto B se conoce que sufrió una intoxicación por *Salmonella*, en el sujeto A el proceso podría estar asociado a un aumento puntual de la abundancia de *Escherichia* durante el periodo de disbiosis. Esta hipótesis se ve reforzada ante el hecho de que en las redes SparCC, *Escherichia* adquiriera una mayor grado o *degree* dentro de las redes.

Los *phyla* Proteobacteria y Actinobacteria son, después de Firmicutes y Bacteroidetes, de los grupos más abundantes e importantes en la microbiota intestinal humana. Se ha documentado que un incremento de proteobacterias está asociado a la disbiosis y al decremento en la abundancia de Firmicutes en individuos con enfermedad inflamatoria intestinal [127], un fenómeno que coincide con las observaciones de los sujetos A y B, donde ambos estuvieron expuestos a un proceso inflamatorio. También se ha asociado que el incremento de Proteobacteria está correlacionado con el decremento de la abundancia de Actinobacteria [129], como se pudo observar en las muestras del sujeto A. También podemos destacar que la alteración en la abundancia de bacterias como *Bifidobacterium* o *Coprococcus* en ambos sujetos durante el periodo de disbiosis, podría estar asociado al desarrollo de un proceso inflamatorio derivado de la infección, pues ambos géneros están asociados a la producción de ácidos grasos de cadena corta, mismos que juegan un papel importante en la regulación de la actividad inflamatoria [130].

Si bien, en las redes SparCC la evidencia no fue concluyente como para asegurar un cambio significativo en la arquitectura de las redes, y por lo tanto, de las comunidades de bacterias encontradas, el hecho de que tanto las abundancias como la arquitectura de las redes ARACNe adoptaran configuraciones similares a las encontradas en el periodo previo a la infección, sugiere que el sistema en su conjunto, a pesar de verse alterado durante periodos de infección, es robusto ante perturbaciones. Cabe destacar que el hecho de que las comunidades de las redes SparCC, ricas en Firmicutes y Proteobacteria, se hayan mantenido en el tiempo a pesar de las alteraciones en las abundancias derivadas de la disbiosis en los sujetos, refuerza la hipótesis de que el sistema parece ser robusto ante perturbaciones, y que ambos *phyla* están ampliamente relacionadas entre sí y juegan un papel importante en el mantenimiento de la

microbiota, aunque se requiere de un análisis más profundo para sustentar esta aseveración.

6.3 Conclusiones generales

El uso de redes multicapa se postula como una herramienta útil en el análisis de la microbiota debido a que permite comparar las propiedades y la arquitectura de distintas redes, ya sea que estas representan segmentos en una escala temporal, como vimos con la evolución de la microbiota a lo largo de un proceso infeccioso, o el estado de la microbiota bajo dos regímenes distintos, como pudimos estudiar cuando comparamos la microbiota de dos organismos estrechamente relacionados en la cadena trófica. Esto nos permite identificar y estudiar patrones que no serían posibles de analizar limitándose a la implementación del enfoque tradicional de la teoría de redes.

Sin embargo, cabe destacar que el hecho de que existan importantes discrepancias en cuanto a la inferencia de redes a partir del uso de distintos algoritmos, indica que si bien las redes de coabundancia pueden ser útiles como herramientas para analizar la posible existencia de interacciones entre microorganismos, sus resultados deben ser tomados con cautela. A lo largo del presente trabajo se emplearon datos de abundancias de los distintos OTUs encontrados en las muestras, lo que plantea importantes limitaciones en cuanto a conocer la naturaleza de las interacciones entre los organismos y su potencial funcionalidad dentro del sistema. Sin embargo, se requiere combinar estas técnicas con otras herramientas, como puede ser el uso de series temporales o de redes de redes que combinan los datos existentes con información sobre los metabolitos secundarios producidos por los microorganismos en cuestión, permitiría ampliar nuestra comprensión, no solo sobre la forma en que los microorganismos interactúan entre sí, sino también sobre cómo afectan a los tejidos del hospedero.

Referencias

- [1] S. E. Schaeffer, "Graph clustering," *Computer science review*, vol. 1, no. 1, pp. 27–64, 2007.
- [2] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [3] F. V. Farahani, W. Karwowski, and N. R. Lighthall, "Application of graph theory for identifying connectivity patterns in human brain networks: A systematic review," *frontiers in Neuroscience*, vol. 13, p. 585, 2019.
- [4] G. Bianconi, *Multilayer networks: Structure and function*. Oxford university press, 2018.

- [5] J. F. Padgett and C. K. Ansell, "Robust action and the rise of the medici, 1400-1434," *American journal of sociology*, vol. 98, no. 6, pp. 1259–1319, 1993.
- [6] E. Cozzo, G. F. de Arruda, F. A. Rodrigues, and Y. Moreno, "Multilayer networks: Metrics and spectral properties," in *Interconnected networks*, Springer, pp. 17–35, 2016.
- [7] E. Cozzo *et al.*, "Clustering coefficients in multiplex networks," *arXiv preprint arXiv:1307.6780*, 2013.
- [8] A. Aleta, S. Meloni, and Y. Moreno, "A multilayer perspective for the analysis of urban transportation systems," *Scientific reports*, vol. 7, no. 1, pp. 1–9, 2017.
- [9] J. Sanz, C.-Y. Xia, S. Meloni, and Y. Moreno, "Dynamics of interacting diseases," *Physical Review X*, vol. 4, no. 4, p. 041005, 2014.
- [10] Y. V. Sun and Y.-J. Hu, "Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases," *Advances in genetics*, vol. 93, pp. 147–190, 2016.
- [11] M. Schleuning *et al.*, "Trait-based assessments of climate-change impacts on interacting species," *Trends in Ecology & Evolution*, vol. 35, no. 4, pp. 319–328, 2020.
- [12] N. Villa-Vialaneix-nathalie, "An introduction to network inference and mining". 2010.
- [13] P. Langfelder and S. Horvath, "WGCNA: An r package for weighted correlation network analysis," *BMC bioinformatics*, vol. 9, no. 1, pp. 1–13, 2008.
- [14] P. E. Meyer, F. Lafitte, and G. Bontempi, "Minet: AR/bioconductor package for inferring large transcriptional networks using mutual information," *BMC bioinformatics*, vol. 9, no. 1, pp. 1–10, 2008.
- [15] J. Friedman and E. J. Alm, "Inferring correlation networks from genomic survey data," 2012.
- [16] Z. D. Kurtz, C. L. Müller, E. R. Miraldi, D. R. Littman, M. J. Blaser, and R. A. Bonneau, "Sparse and compositionally robust inference of microbial ecological networks," *PLoS computational biology*, vol. 11, no. 5, p. e1004226, 2015.
- [17] H. Hirano and K. Takemoto, "Difficulty in inferring microbial community structure based on co-occurrence network approaches," *BMC bioinformatics*, vol. 20, no. 1, pp. 1–14, 2019.
- [18] Y. Fan and O. Pedersen, "Gut microbiota in human metabolic health and disease," *Nature Reviews Microbiology*, vol. 19, no. 1, pp. 55–71, 2021.
- [19] F. Del Chierico *et al.*, "Gut microbiota profiling of pediatric nonalcoholic fatty liver disease and obese patients unveiled by an integrated meta-omics-based approach," *Hepatology*, vol. 65, no. 2, pp. 451–464, 2017.

- [20] M. I. Smith *et al.*, “Gut microbiomes of malawian twin pairs discordant for kwashiorkor,” *Science*, vol. 339, no. 6119, pp. 548–554, 2013.
- [21] H. Zheng, H. Wang, R. J. Dewhurst, and R. Roehe, “Improving the inference of co-occurrence networks in the bovine rumen microbiome,” *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 17, no. 3, pp. 858–867, 2018.
- [22] L. Viniegra Velázquez, “El reduccionismo científico y el control de las conciencias: Parte i,” *Boletín Médico del Hospital Infantil de México*, vol. 71, no. 4, pp. 252–257, 2014.
- [23] P. W. Anderson, “More is different,” *Science*, vol. 177, no. 4047, pp. 393–396, 1972.
- [24] B. Falkenburg and M. Morrison, “Why more is different,” *Philosophical Issues in Condensed Matter Physics and Complex Systems*. Springer, Berlin, Heidelberg, 2015.
- [25] M. Gardner, “The fantastic combinations of jhon conway’s new solitaire game’Life,” *Sc. Am.*, vol. 223, pp. 20–123, 1970.
- [26] J. Von Neuman *et al.*, “Theory of self-reproducing automata,” 1966.
- [27] O. Mason and M. Verwoerd, “Graph theory and networks in biology,” *IET systems biology*, vol. 1, no. 2, pp. 89–119, 2007.
- [28] J. Stelling, U. Sauer, Z. Szallasi, F. J. Doyle III, and J. Doyle, “Robustness of cellular functions,” *Cell*, vol. 118, no. 6, pp. 675–685, 2004.
- [29] L. A. Zuzek, H. E. Stanley, and L. A. Braunstein, “Epidemic model with isolation in multilayer networks,” *Scientific reports*, vol. 5, no. 1, pp. 1–7, 2015.
- [30] A. Aleta and Y. Moreno, “Multilayer networks in a nutshell,” *Annual Review of Condensed Matter Physics*, vol. 10, pp. 45–62, 2019.
- [31] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.
- [32] M. E. Newman, A.-L. E. Barabási, and D. J. Watts, *The structure and dynamics of networks*. Princeton university press, 2006.
- [33] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, “Controllability of complex networks,” *nature*, vol. 473, no. 7346, pp. 167–173, 2011.
- [34] Y. Wang, E. Ghumare, R. Vandenberghe, and P. Dupont, “Comparison of different generalizations of clustering coefficient and local efficiency for weighted undirected graphs,” *Neural computation*, vol. 29, no. 2, pp. 313–331, 2017.
- [35] M. Newman and G. Barkema, “Oxford university press; new york” *Networks: An Introduction*. [Google Scholar]. 2010.
- [36] A. B. Das, “Small-world networks of prognostic genes associated with lung adenocarcinoma development,” *Genomics*, vol. 112, no. 6, pp. 4078–4088, 2020.

- [37] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.
- [38] R. Kanawati, "Multiplex network mining: A brief survey." *IEEE Intell. Informatics Bull.*, vol. 16, no. 1, pp. 24–27, 2015.
- [39] P. Holme and J. Saramäki, "Temporal networks," *Physics reports*, vol. 519, no. 3, pp. 97–125, 2012.
- [40] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical processes on complex networks*. Cambridge university press, 2008.
- [41] D. Y. Kenett, M. Perc, and S. Boccaletti, "Networks of networks—an introduction," *Chaos, Solitons & Fractals*, vol. 80, pp. 1–6, 2015.
- [42] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas, "Diffusion dynamics on multiplex networks," *Physical review letters*, vol. 110, no. 2, p. 028701, 2013.
- [43] J. Iacovacci, C. Rahmede, A. Arenas, and G. Bianconi, "Functional multiplex pagerank," *EPL (Europhysics Letters)*, vol. 116, no. 2, p. 28004, 2016.
- [44] A. Lancichinetti and S. Fortunato, "Consensus clustering in complex networks," *Scientific reports*, vol. 2, no. 1, pp. 1–7, 2012.
- [45] J. F. Donges, H. C. Schultz, N. Marwan, Y. Zou, and J. Kurths, "Investigating the topology of interacting networks," *The European Physical Journal B*, vol. 84, no. 4, pp. 635–651, 2011.
- [46] Y.-Y. Liu and A.-L. Barabási, "Control principles of complex systems," *Reviews of Modern Physics*, vol. 88, no. 3, p. 035006, 2016.
- [47] E. Lazega and P. E. Pattison, "Multiplexity, generalized exchange and cooperation in organizations: A case study," *Social networks*, vol. 21, no. 1, pp. 67–90, 1999.
- [48] B. Bentley *et al.*, "The multilayer connectome of caenorhabditis elegans," *PLoS computational biology*, vol. 12, no. 12, p. e1005283, 2016.
- [49] C. I. Del Genio, J. Gómez-Gardeñes, I. Bonamassa, and S. Boccaletti, "Synchronization in networks with multiple interaction layers," *Science Advances*, vol. 2, no. 11, p. e1601679, 2016.
- [50] O. Schweiger, J. Settele, O. Kudrna, S. Klotz, and I. Kühn, "Climate change can cause spatial mismatch of trophically interacting species," *Ecology*, vol. 89, no. 12, pp. 3472–3479, 2008.
- [51] A. J. Butte and I. S. Kohane, "Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements," in *Biocomputing 2000*, World Scientific, pp. 418–429. 1999.

- [52] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," *Statistical applications in genetics and molecular biology*, vol. 4, no. 1, 2005.
- [53] C. Uhler, "Gaussian graphical models: An algebraic and geometric perspective," *arXiv preprint arXiv:1707.04345*, 2017.
- [54] A. A. Margolin *et al.*, "ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context," in *BMC bioinformatics*, vol. 7, pp. 1–15. 2006.
- [55] E. Z. Chen and H. Li, "A two-part mixed-effects model for analyzing longitudinal microbiome compositional data," *Bioinformatics*, vol. 32, no. 17, pp. 2611–2617, 2016.
- [56] M. Layeghifard, D. M. Hwang, and D. S. Guttman, "Disentangling interactions in the microbiome: A network perspective," *Trends in microbiology*, vol. 25, no. 3, pp. 217–228, 2017.
- [57] M. Layeghifard, D. M. Hwang, and D. S. Guttman, "Constructing and analyzing microbiome networks in R," in *Microbiome analysis*, Springer, pp. 243–266. 2018.
- [58] A. E. Gamiño-Arroyo, M. P. Barrios-Ceballos, L. P. C. de la Peña, F. Anaya-Velázquez, and F. Padilla-Vaca, "Flora normal, probióticos y salud humana," *Acta universitaria*, vol. 15, no. 3, pp. 34–40, 2005.
- [59] D. Rothschild *et al.*, "Environment dominates over host genetics in shaping human gut microbiota," *Nature*, vol. 555, no. 7695, pp. 210–215, 2018.
- [60] T. Korem *et al.*, "Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples," *Science*, vol. 349, no. 6252, pp. 1101–1106, 2015.
- [61] G. Falony, S. Vieira-Silva, and J. Raes, "Richness and ecosystem development across faecal snapshots of the gut microbiota," *Nature microbiology*, vol. 3, no. 5, pp. 526–528, 2018.
- [62] T. Yatsunenko *et al.*, "Human gut microbiome viewed across age and geography," *nature*, vol. 486, no. 7402, pp. 222–227, 2012.
- [63] J. Qin *et al.*, "A human gut microbial gene catalogue established by metagenomic sequencing," *nature*, vol. 464, no. 7285, pp. 59–65, 2010.
- [64] F. Bäckhed *et al.*, "The gut microbiota as an environmental factor that regulates fat storage," *Proceedings of the national academy of sciences*, vol. 101, no. 44, pp. 15718–15723, 2004.
- [65] L. M. Cox and M. J. Blaser, "Antibiotics in early life and obesity," *Nature Reviews Endocrinology*, vol. 11, no. 3, pp. 182–190, 2015.
- [66] A. Górski *et al.*, "Perspectives of phage therapy in non-bacterial infections," *Frontiers in Microbiology*, vol. 9, p. 3306, 2019.

- [67] J. C. Clemente *et al.*, "The microbiome of uncontacted amerindians," *Science advances*, vol. 1, no. 3, p. e1500183, 2015.
- [68] S. A. Smits *et al.*, "Seasonal cycling in the gut microbiome of the hadza hunter-gatherers of tanzania," *Science*, vol. 357, no. 6353, pp. 802–806, 2017.
- [69] A. Flint, A. Raben, J. Rehfeld, J. Holst, and A. Astrup, "The effect of glucagon-like peptide-1 on energy expenditure and substrate metabolism in humans," *International journal of obesity*, vol. 24, no. 3, pp. 288–298, 2000.
- [70] K. M. Maslowski *et al.*, "Regulation of inflammatory responses by gut microbiota and chemoattractant receptor gpr43," *Nature*, vol. 461, no. 7268, pp. 1282–1286, 2009.
- [71] S. Macfarlane and G. T. Macfarlane, "Composition and metabolic activities of bacterial biofilms colonizing food residues in the human gut," *Applied and environmental microbiology*, vol. 72, no. 9, pp. 6204–6211, 2006.
- [72] W. R. Russell *et al.*, "Major phenylpropanoid-derived metabolites in the human gut can arise from microbial fermentation of protein," *Molecular nutrition & food research*, vol. 57, no. 3, pp. 523–535, 2013.
- [73] V. D. de Mello *et al.*, "Indolepropionic acid and novel lipid metabolites are associated with a lower risk of type 2 diabetes in the finnish diabetes prevention study," *Scientific reports*, vol. 7, no. 1, pp. 1–12, 2017.
- [74] S. Tims *et al.*, "Microbiota conservation and bmi signatures in adult monozygotic twins," *The ISME journal*, vol. 7, no. 4, pp. 707–717, 2013.
- [75] U. Gophna, T. Konikoff, and H. B. Nielsen, "Oscillospira and related bacteria—from metagenomic species to metabolic features," *Environmental microbiology*, vol. 19, no. 3, pp. 835–841, 2017.
- [76] T. L. Miller, M. Wolin, E. C. de Macario, and A. Macario, "Isolation of methanobrevibacter smithii from human feces," *Applied and environmental microbiology*, vol. 43, no. 1, pp. 227–232, 1982.
- [77] V. K. Ridaura *et al.*, "Gut microbiota from twins discordant for obesity modulate metabolism in mice," *Science*, vol. 341, no. 6150, 2013.
- [78] C. A. Thaiss *et al.*, "Hyperglycemia drives intestinal barrier dysfunction and risk for enteric infection," *Science*, vol. 359, no. 6382, pp. 1376–1383, 2018.
- [79] K. H. Allin *et al.*, "Aberrant intestinal microbiota in individuals with prediabetes," *Diabetologia*, vol. 61, no. 4, pp. 810–820, 2018.
- [80] T. Bryrup *et al.*, "Metformin-induced changes of the gut microbiota in healthy young men: Results of a non-blinded, one-armed intervention study," *Diabetologia*, vol. 62, no. 6, pp. 1024–1035, 2019.

[81] A. Adeshirlarijaney, J. Zou, H. Q. Tran, B. Chassaing, and A. T. Gewirtz, "Amelioration of metabolic syndrome by metformin associates with reduced indices of low-grade inflammation independently of the gut microbiota," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 317, no. 6, pp. E1121–E1130, 2019.

[82] Y. Fan *et al.*, "Comprehensive metabolomic characterization of coronary artery diseases," *Journal of the American College of Cardiology*, vol. 68, no. 12, pp. 1281–1293, 2016.

[83] E. D. Michos, J. W. McEvoy, and R. S. Blumenthal, "Lipid management for the prevention of atherosclerotic cardiovascular disease," *New England Journal of Medicine*, vol. 381, no. 16, pp. 1557–1567, 2019.

[84] H. Sokol *et al.*, "Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of crohn disease patients," *Proceedings of the National Academy of Sciences*, vol. 105, no. 43, pp. 16731–16736, 2008.

[85] X. Cui *et al.*, "Metagenomic and metabolomic analyses unveil dysbiosis of gut microbiota in chronic heart failure patients," *Scientific reports*, vol. 8, no. 1, pp. 1–15, 2018.

[86] W. Zhu *et al.*, "Gut microbial metabolite tmao enhances platelet hyperreactivity and thrombosis risk," *Cell*, vol. 165, no. 1, pp. 111–124, 2016.

[87] V. Senthong *et al.*, "Intestinal microbiota-generated metabolite trimethylamine-n-oxide and 5-year mortality risk in stable coronary artery disease: The contributory role of intestinal microbiota in a courage-like patient cohort," *Journal of the American Heart Association*, vol. 5, no. 6, p. e002816, 2016.

[88] H. L. Collins *et al.*, "L-carnitine intake and high trimethylamine n-oxide plasma levels correlate with low aortic lesions in apoe-/- transgenic mice expressing cetp," *Atherosclerosis*, vol. 244, pp. 29–37, 2016.

[89] R. Rao, A. Seth, and P. Sheth, "Recent advances in alcoholic liver disease i. Role of intestinal permeability and endotoxemia in alcoholic liver disease," *American Journal of Physiology-Gastrointestinal and Liver Physiology*, vol. 286, no. 6, pp. G881–G884, 2004.

[90] S. E. Zytynska, "Embracing the complexity of plant–microbe–insect interactions under a changing climate for sustainable agriculture," *Current opinion in insect science*, vol. 44, pp. 89–94, 2021.

[91] P. Trivedi, J. E. Leach, S. G. Tringe, T. Sa, and B. K. Singh, "Plant–microbiome interactions: From community assembly to plant health," *Nature reviews microbiology*, vol. 18, no. 11, pp. 607–621, 2020.

[92] J. Vorholt, "Microbial life in the phyllosphere. Nat publ gr 10: 828–840." 2012.

[93] H. Pulido, K. E. Mauck, C. M. De Moraes, and M. C. Mescher, "Combined effects of mutualistic rhizobacteria counteract virus-induced suppression of indirect plant

defences in soya bean,” *Proceedings of the Royal Society B*, vol. 286, no. 1903, p. 20190211, 2019.

[94] L. del Rosario Cappellari, J. Chiappero, T. B. Palermo, W. Giordano, and E. Banchio, “Impact of soil rhizobacteria inoculation and leaf-chewing insect herbivory on mentha piperita leaf secondary metabolites,” *Journal of Chemical Ecology*, vol. 46, no. 7, pp. 619–630, 2020.

[95] K. Bell, N. Naranjo-Guevara, R. C. dos Santos, R. Meadow, and J. Bento, “Predatory earwigs are attracted by herbivore-induced plant volatiles linked with plant growth-promoting rhizobacteria,” *Insects*, vol. 11, no. 5, p. 271, 2020.

[96] C. H. Haney *et al.*, “Rhizosphere-associated pseudomonas induce systemic resistance to herbivores at the cost of susceptibility to bacterial pathogens,” *Molecular Ecology*, vol. 27, no. 8, pp. 1833–1847, 2018.

[97] R. Guimera and L. A. N. Amaral, “Cartography of complex networks: Modules and universal roles,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 2, p. P02001, 2005.

[98] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, “Detecting rich-club ordering in complex networks,” *Nature physics*, vol. 2, no. 2, pp. 110–115, 2006.

[99] S. N. Baldassano and D. S. Bassett, “Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease,” *Scientific reports*, vol. 6, no. 1, pp. 1–14, 2016.

[100] Z. Mayoral-Peña, R. Álvarez-Martínez, J. Fornoni, and E. Garrido, “The extended microbiota: How microbes shape plant-insect interactions,” in *Evolutionary ecology of plant-herbivore interaction*, Springer, pp. 135–146. 2020.

[101] K. Faust, F. Bauchinger, S. de Buyl, and L. Lahti, *Seqtime: Time series analysis of sequencing data*. 2021. Available: <https://github.com/hallucigenia-sparsa/seqtime>

[102] F. David Materna, “Host lifestyle affects human microbiota on daily timescales,” *Genome Biology*, vol. 7, 2014.

[103] G. Csardi and T. Nepusz, “The igraph software package for complex network research,” *InterJournal*, vol. Complex Systems, p. 1695, 2006, Available: <https://igraph.org>

[104] P. J. McMurdie and S. Holmes, “Phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data,” *PLoS ONE*, vol. 8, no. 4, p. e61217, 2013.

[105] Z. Kurtz, C. Mueller, E. Miraldi, and R. Bonneau, *SpiecEasi: Sparse inverse covariance for ecological statistical inference*. 2021.

- [106] P. E. Meyer, F. Lafitte, and G. Bontempi, "MINET: An open source r/bioconductor package for mutual information based network inference," *BMC Bioinformatics*, vol. 9, 2008.
- [107] M. De Domenico, *MuxViz: MuxViz - visualization and analysis of multilayer networks*. 2021.
- [108] M. Magnani, L. Rossi, and D. Vega, "Analysis of multiplex social networks with R," *Journal of Statistical Software*, vol. 98, no. 8, pp. 1–30, 2021, doi: [10.18637/jss.v098.i08](https://doi.org/10.18637/jss.v098.i08).
- [109] T. Wei and V. Simko, *R package 'corrplot': Visualization of a correlation matrix*. 2021.
- [110] R. Kindt and R. Coe, *Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies*. Nairobi (Kenya): World Agroforestry Centre (ICRAF), 2005.
- [111] B. Tóthmérész, "Comparison of different methods for diversity ordering," *Journal of vegetation Science*, vol. 6, no. 2, pp. 283–290, 1995.
- [112] D. Combe, C. Llargeron, M. Géry, and E. Egyed-Zsigmond, "I-louvain: An attributed graph clustering method," in *International symposium on intelligent data analysis*, pp. 181–192. 2015.
- [113] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Generalized louvain method for community detection in large networks," in *2011 11th international conference on intelligent systems design and applications*, pp. 88–93. 2011.
- [114] G. Klein and J. E. Aronson, "Optimal clustering: A model and method," *Naval Research Logistics (NRL)*, vol. 38, no. 3, pp. 447–461, 1991.
- [115] L. A. Kazakovtsev and A. N. Antamoshkin, "Genetic algorithm with fast greedy heuristic for clustering and location problems," *Informatica*, vol. 38, no. 3, 2014.
- [116] K. Y. Yeung and W. L. Ruzzo, "Details of the adjusted rand index and clustering algorithms, supplement to the paper an empirical study on principal component analysis for clustering gene expression data," *Bioinformatics*, vol. 17, no. 9, pp. 763–774, 2001.
- [117] L. Ana and A. K. Jain, "Robust data clustering," in *2003 ieee computer society conference on computer vision and pattern recognition, 2003. Proceedings.*, vol. 2, pp. II–II. 2003.
- [118] J. K. Goldberg, "Towards understanding the mechanisms maintaining a stable trichome dimorphism in the sacred datura plant (*datura wrightii*)," PhD thesis, Indiana University, 2020.

- [119] M. del Pilar Martínez-Diz, M. Andrés-Sodupe, R. Bujanda, E. Díaz-Losada, A. Eichmeier, and D. Gramaje, "Soil-plant compartments affect fungal microbiome diversity and composition in grapevine," *Fungal Ecology*, vol. 41, pp. 234–244, 2019.
- [120] C. Fausto *et al.*, "Olive orchard microbiome: Characterisation of bacterial communities in soil-plant compartments and their comparison between sustainable and conventional soil management systems," *Plant ecology & diversity*, vol. 11, nos. 5-6, pp. 597–610, 2018.
- [121] K. Visakorpi, S. Gripenberg, Y. Malhi, and T. Riutta, "Does insect herbivory suppress ecosystem productivity? Evidence from a temperate woodland," *bioRxiv*, 2021.
- [122] F. Renoz *et al.*, "Evidence for gut-associated *Serratia symbiotica* in wild aphids and ants provides new perspectives on the evolution of bacterial mutualism in insects," *Microbial ecology*, vol. 78, no. 1, pp. 159–169, 2019.
- [123] E. K. Costello, C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight, "Bacterial community variation in human body habitats across space and time," *science*, vol. 326, no. 5960, pp. 1694–1697, 2009.
- [124] J. G. Caporaso *et al.*, "Moving pictures of the human microbiome," *Genome biology*, vol. 12, no. 5, pp. 1–8, 2011.
- [125] L. Dethlefsen and D. A. Relman, "Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation," *Proceedings of the National Academy of Sciences*, vol. 108, no. Supplement 1, pp. 4554–4561, 2011.
- [126] L. A. David *et al.*, "Host lifestyle affects human microbiota on daily timescales," *Genome biology*, vol. 15, no. 7, pp. 1–15, 2014.
- [127] F. Magne *et al.*, "The firmicutes/bacteroidetes ratio: A relevant marker of gut dysbiosis in obese patients?" *Nutrients*, vol. 12, no. 5, p. 1474, 2020.
- [128] C. M. dos S. P. Indiani, K. F. Rizzardi, P. M. Castelo, L. F. C. Ferraz, M. Darrieux, and T. M. Parisotto, "Childhood obesity and firmicutes/bacteroidetes ratio in the gut microbiota: A systematic review," *Childhood obesity*, vol. 14, no. 8, pp. 501–509, 2018.
- [129] S. Stojanov, A. Berlec, and B. Štrukelj, "The influence of probiotics on the firmicutes/bacteroidetes ratio in the treatment of obesity and inflammatory bowel disease," *Microorganisms*, vol. 8, no. 11, p. 1715, 2020.
- [130] S. Lee, K. I. Keirse, R. Kirkland, Z. I. Grunewald, J. G. Fischer, and C. B. de La Serre, "Blueberry supplementation influences the gut microbiota, inflammation, and insulin resistance in high-fat-diet-fed rats," *The Journal of nutrition*, vol. 148, no. 2, pp. 209–219, 2018.