



UNIVERSIDAD NACIONAL DE COLOMBIA

Una mirada a la evolución de las vías metabólicas desde la perspectiva de las redes complejas

Francisco Javier Rodríguez Lanza

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Física
Bogotá, Colombia
2020

Una mirada a la evolución de las vías metabólicas desde la perspectiva de las redes complejas

Francisco Javier Rodríguez Lanza

Director:

Ph.D. Rafael Germán Hurtado Heredia

Codirectora:

Ph.D. Liliana Lopez Kleine

Grupo de Investigación:

Sociofísica y Econofísica / Bioinformática y Biología de Sistemas

Universidad Nacional de Colombia

Facultad de Ciencias, Departamento de Física

Bogotá, Colombia

2020

Resumen

En este trabajo de grado, se modelan las redes metabólicas de 13 organismos con trayectorias evolutiva distintas, por medio de dos tipos distintos de redes, las redes centradas en enzimas y las redes bipartitas de enzimas y compuestos. Se hace un análisis comparativo y exploratorio de sus propiedades estructurales con el objetivo de encontrar correlaciones con sus distancias evolutivas. Se encontró que los metabolismos de los organismos estudiados son estructuralmente diferentes y que su grado de divergencia, en algunas variables topológicas y estadísticas como el grado promedio y la diferencia de grado, se corresponden estrechamente con las distancias evolutivas entre ellos. También se obtuvieron distribuciones de grado exponenciales en las redes de enzimas. Según estos resultados las redes metabólicas contienen información filogenética relevante, y en su proceso evolutivo probablemente se involucran mecanismos de crecimiento que favorecen la diversidad estructural.

Palabras clave: Redes metabólicas, Evolución biológica, Redes complejas.

Abstract

In this bachelor's thesis, the metabolic network of 13 organisms with different evolutionary trajectories are modeled using two different types of networks, the ones that are centered on enzymes and bipartite networks of enzymes and compounds. A comparative and exploratory analysis is made over its structural properties in order to find correlations with its evolutionary distances. It was found that the metabolisms of those organisms are structurally different and their divergence degree, in some topological and statistical variables such as the average degree, and the degree difference closely correspond to its evolutive distances. Exponential degree distributions were also obtained, in those enzyme network. According to these results, metabolic networks contain relevant phylogenetic information and in its evolutive process are likely involved growth mechanisms that stimulate structural diversity..

Keywords: Metabolic networks, Evolutionary biology, Complex networks.

Contenido

1	Introducción	2
1.1	Objetivos	3
2	Marco Teórico	4
2.1	Evolución Biología	4
2.1.1	Árbol Filogenético	4
2.1.2	Biología molecular y evolución	5
2.2	Naturaleza compleja	6
2.2.1	Redes	6
2.3	Procesos Metabólicos	7
2.3.1	Vías Metabólicas	8
2.3.2	Redes Metabólicas	9
3	Metodología	11
3.1	Bases de Datos	11
3.2	Teoría de grafos	13
3.2.1	Medidas de grafo	13
3.3	Propiedades estadísticas de las redes complejas	18
3.3.1	Distribuciones de grado	18
3.3.2	Entropía	20
3.3.3	Divergencia de Kullback-Leibler	21
3.4	Análisis comparativo de Redes	21
3.4.1	Análisis de Componente Principales - ACP	22
3.4.2	Asimetría de una distribución.	22
3.4.3	Correlación de Pearson	23
4	Resultados y discusión	24
4.1	Análisis comparativo de redes metabólicas implementando un ACP	24
4.1.1	Discusión	29
4.2	Análisis comparativo del grado nodal de las Redes metabólicas	30
4.2.1	Las tuplas $[k_{out}, k_{in}]$	30
4.2.2	Discusión	33
4.2.3	La diferencia de grado $k_{out} - k_{in}$	35
4.2.4	Discusión	39

4.3	Comparación de las dinámicas de las Redes metabólicas	40
4.3.1	Distribución de grado nodal	40
4.3.2	Discusión	43
5	Conclusiones e hipótesis	45
5.1	Conclusiones	45
5.2	Hipótesis	46
6	Anexos	47
6.1	Prevalencia de las diferentes clases de enzimas en las Redes metabólicas . . .	47
	Bibliografía	50

1 Introducción

Aunque en la actualidad no existe un consenso sobre la definición de la vida y sobre su origen en la Tierra, sí se puede decir que ésta implica una amplia serie de conceptos que la describen y diferencian de otros conglomerados de materia, entre estos los de mayor relevancia son conceptos de organización, complejidad, replicación y evolución [1]. La Biología evolutiva, campo que se encarga del estudio de los orígenes y los cambios a través del tiempo de la vida, ha encontrado que su desarrollo probablemente se dio por la articulación mutua de sistemas como vesículas membranosas, polímeros replicativos y las vías metabólicas. Es por esta razón que estudiar cómo se organiza el metabolismo y cómo evoluciona es un desafío central en la ciencias biológicas para entender la vida en la Tierra.

El metabolismo constituye una red de interacciones catalíticas que regulan diversos procesos biológicos al interior de los seres vivos, es sin duda uno de los procesos más importantes que lleva a cabo la vida y se asume que son resultado de la selección natural actuando a nivel molecular [2]. Una de las teorías más aceptadas sobre el desarrollo del metabolismo sostiene que en los comienzos de la vida, las protocelulas en la sopa prebiótica poseían sistemas metabólicos simples capaces de sintetizar los nutrientes presentes, y que a medida que se iban agotando, los organismos se vieron presionados a evolucionar desarrollando sistemas metabólicos más sofisticados que les permitieran obtener los nutrientes de precursores más simples y abundantes [3].

En la actualidad, en el estudio del sistema metabólico, debido a las serias limitaciones que presentaban su análisis a gran escala desde el enfoque de las rutas metabólicas lineales e independientes, se han introducido técnicas de análisis de redes complejas como un enfoque necesario para entender la complejidad de los procesos bioquímicos que lleva a cabo el metabolismo y permitir su comparación estructural entre diferentes especies. De esta unión nace el concepto de red metabólica en cuyos análisis topológicos resultan diferencias cualitativas y cuantitativas de la organización metabólica que son fundamento de hipótesis sobre las estructuras de las redes metabólicas y la información filogenética que aportan a la clasificación evolutiva de los organismos [4].

1.1. Objetivos

Los objetivos generales y específicos de este trabajo de grado han sido:

Objetivo general

Caracterizar y comparar redes metabólicas de organismos con trayectorias evolutivas distintas.

Objetivos específicos

- Obtener las redes metabólicas de organismos con trayectorias evolutivas distintas.
- Establecer propiedades topológicas y de contenido de información de las redes metabólicas.
- Relacionar las propiedades topológicas de las redes metabólicas con las trayectorias evolutivas de los organismos.

2 Marco Teórico

Esta sección se divide en tres partes, en la primera se tratan de forma introductoria conceptos de evolución biológica y su relación con la Biología molecular, posteriormente se describen los sistemas complejos y las redes, y por ultimo se hace una breve descripción del metabolismos celular.

2.1. Evolución Biología

En 1885 el naturalista británico Charles Darwin publicó “El origen de las especies”, un texto revolucionario en el cual se postula la teoría de la evolución de las especies, la cual se propone como una explicación a la adaptabilidad que presentan las especies al habitat y a la gran diversidad de formas de vida que existen en el planeta, esta también propone la selección natural como el mecanismo que propicia la evolución, el cual consiste en la reproducción diferencial de fenotipos (cualidades morfológicas) de un población en función de ventajas adaptativas que favorecen la supervivencia, y la idea de que todas las formas de vida se encuentran relacionadas por medio de un antecesor común [5]. En la actualidad el estudio de la evolución de las especies es abordado por la Biología evolutiva, la cual comprende el estudio tanto molecular como ecológico de las diferencias y relaciones entre organismos, y de los mecanismos por los cuales se establecen [6]. En la actualidad se ha teorizado sobre diversos mecanismos por los cuales las especies evolucionan, entre ellos está la deriva genética.

Una de las ideas más significativas mencionadas anteriormente, y que es un pilar metodológico de la Biología evolutiva, es que todos las formas de vida se encuentran relacionadas en algún grado por medio de un ancestro común. Estas relaciones evolutivas son observables a través de las características anatómicas, funcionales y genéticas de los organismos, con este idea se investiga los orígenes evolutivos y se estructura la historia de la vida en el planeta, un diagrama de estas relaciones evolutivas entre especies es el árbol de la vida o árbol filogenético.

2.1.1. Árbol Filogenético

Un árbol filogenético, figura [2-1], es un representación arborescente donde se muestran la relaciones evolutivas entre especies o grupos de especies que llamamos taxones. Estos arboles no son una representación absoluta de los procesos evolutivos, mas sin embargo

reúne información fósil, anatómica, funcional y molecular de los organismos por medio del cual se establecen proximidades y divergencias [7].

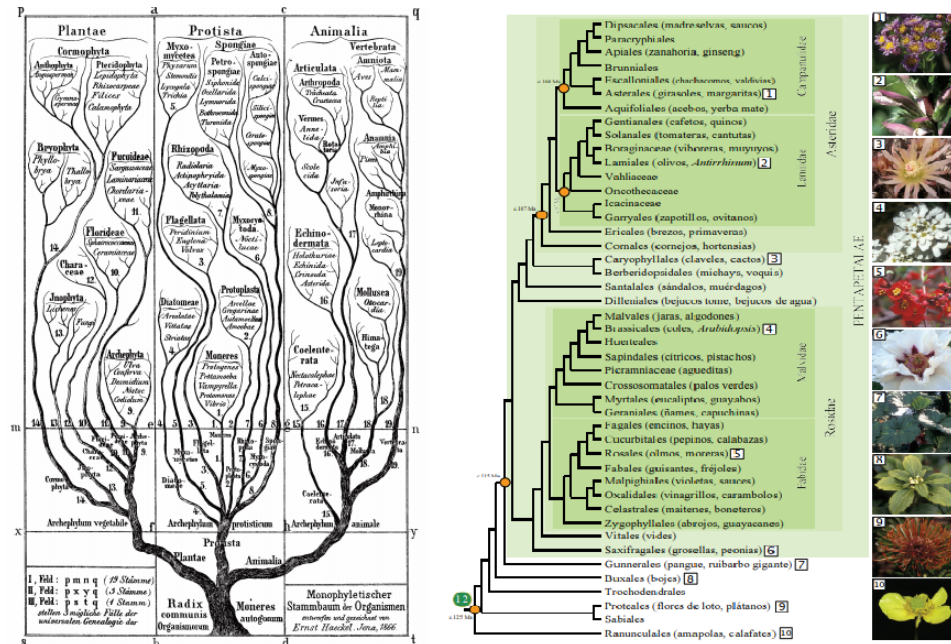


Figura 2-1: En la imagen de la izquierda se muestra el primer árbol filogenético universal, publicado por Haeckel en 1866 y en la imagen de la derecha se muestra el árbol filogenético de las Eudicotiledóneas, con 37 de los taxones mas importantes. Tomado de VARGAS, Pablo; ZARDOYA, Rafael. El árbol de la vida: sistemática y evolución de los seres vivos. Vargas, Pablo, 2012.

Un cronograma es un tipo de árbol filogenético en el cual las longitudes de las ramas son proporcionales a los tiempos evolutivos, en esta representaciones se mide el tiempo desde un taxón de referencia a las ramificación próximas [7]. En este sentido, hablamos de distancias evolutivas como las longitudes de las ramas que conectan dos o mas organismos.

2.1.2. Biología molecular y evolución

Aunque las características anatómicas visibles que un organismo presenta, que llamaremos fenotipo, le permitieron a Darwin desarrollar y darle consistencia a su teoría, mas adelante se descubrió que habían características químicas y de funcionamiento interno de los organismos, que llamaremos genotipo, en las cuales prevalecían estas diferencias y similitudes, y que resolvían problemas de clasificación evolutiva que no se podía por medio del estudio del fenotipo.

Actualmente la Biología molecular ha descubierto y descrito cientos de procesos bioquímicos entre los cuales se encuentran los mecanismos de regulación de la coagulación sanguínea, la

expresión genética y los de regulación de las vías metabólicas, muy probablemente la vida surgió de la interacción de procesos bioquímicos precursores mas simples [8]. Por medio del estudio molecular de los organismos, fue posible también conocer los mecanismos químicos de la herencia y las mutaciones que permiten el desarrollo de las diferencias que propician la evolución.

2.2. Naturaleza compleja

En la actualidad los sistemas complejos están presentes dentro del vocabulario de muchas áreas del conocimiento, y conforme pasa el tiempo toma cada vez mas relevancia en los diferentes ámbitos de la ciencia y de la cotidianidad, este hecho hace que dar una definición única y simple de lo que son este tipo de sistemas no sea fácil [9]. Sin embargo hay algunas características universales que les podemos asociar, como el hecho de ser sistemas compuestos de muchas partes que interactúan entre si, y que su respuesta a cambios en algunas de estas partes es altamente no lineal [10]. Para los objetivos de esta investigación, nos referimos por este termino a aquellos sistemas que no en virtud del número de elementos que le constituyen, sino de las interacciones entre estos, presentan comportamientos emergentes o colectivos no triviales, y cuyo entendimiento escapa a la lógica de los estudios reduccionistas, en los cuales se aproxima al sistema mediante la adición de los comportamientos y mecanismos de sus elementos constituyentes.[11]

Las interacciones entre los elementos de un sistema complejo contienen información adicional, que debe tenerse en cuenta si se quiere hacer una descripción adecuada de su funcionamiento. En la naturaleza encontramos una gran cantidad de estos sistemas, un ejemplo es el sistema nervioso humano, constituido por millones de neuronas interconectadas, que en su conjunto dan lugar a propiedades cognitivas como la memoria, cuya comprensión escapa al estudio del funcionamiento de la neurona, otro ejemplo muy habitual y fascinante es la colonia de hormigas en la cual aun sin tener un control centralizado las hormigas se organizan para construir estructuras tan complejas como un hormiguero, este ejemplo saca a la luz una de las propiedades mas interesantes descubiertas en este tipo de sistemas que es la autoorganización.

2.2.1. Redes

Como lo mencionamos anteriormente, en un sistema complejo son las interacciones entre sus partes las que determinan la dinámica y el funcionamiento colectivo del sistema, una forma de modelar estas interacciones es mediante el uso de redes y grafos. Las redes contienen informacion sobre la naturaleza de los elementos y sus relaciones, los grafos son estructuras matemáticas en la cuales se representan dichas interacciones entre los elementos mediante una colección de nodos y aristas. Estas representaciones, además de permitirnos visualizar el

sistema en su conjunto, nos permite analizarlo cuantitativamente con una amplia variedad de herramientas de la teoría de grafos.

Podemos clasificar las redes en dos conjuntos, dirigidas y no dirigidas [2-2]. Los grafos no dirigidos representan sistemas complejos en los cuales las interacciones entre sus elementos son simétricas y no apuntan en ningún sentido en particular, “un ejemplo de este tipo de sistemas es la red de contactos sexuales en la cual se tiene que si una persona A tuvo relaciones con una persona B, entonces B evidentemente tuvo relaciones con A” [10]. Por otra parte muchos sistemas complejos tienen interacciones direccionadas representadas por redes dirigidas, como por ejemplo la red de transmisión del SARS-CoV-2, en el cual una persona A infecta a otra persona B, pero no es posible la persona B contagie entonces a la persona A.

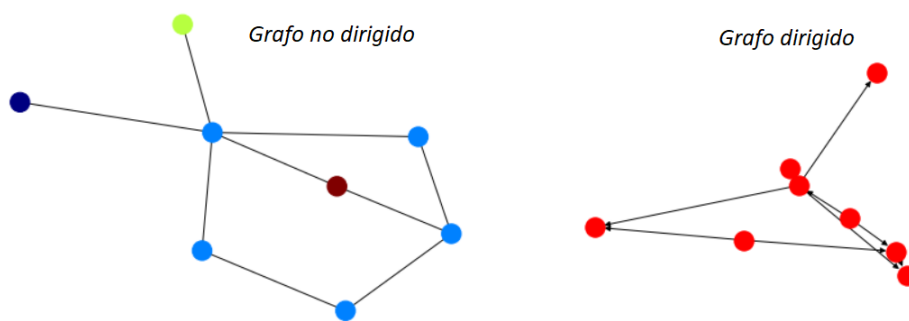


Figura 2-2: Representación gráfica de redes dirigidas y no dirigidas

En las redes dirigidas podemos distinguir tres tipos diferentes de nodos dependiendo de su conectividad:

- Receptor: Si el nodo n_i cumple que $k_{in}^i > 0$ y $k_{out}^i = 0$
- Transmisor: Si el nodo n_i cumple que $k_{in}^i = 0$ y $k_{out}^i > 0$
- Ordinario: Si el nodo n_i cumple que $k_{in}^i > 0$ y $k_{out}^i > 0$

Donde k_{in} y k_{out} son el grado de entrada y de salida que trataremos mas adelante.

2.3. Procesos Metabólicos

El metabolismo celular es uno de los procesos fundamentales de la vida. Está compuesto de moléculas orgánicas, usualmente de naturaleza proteica, llamadas enzimas Figura[2-3], capaces de catalizar, es decir propiciar, reacciones químicas. Las enzimas son las encargadas de transformar los alimentos que recibe la célula en energía o compuestos estructurales

necesarios para soportar sus funciones vitales, siendo uno de los procesos mas complejos y organizados que se lleva acabo dentro de la célula, debido a que la obtención de un componente bioquímico vital debe ser obtenido a partir de un alimento o sustrato externo cuya transformación depende usualmente de un conjunto de reacciones químicas, llevadas acabo de forma organizada y una a una por las enzimas que conforman el metabolismo [12].

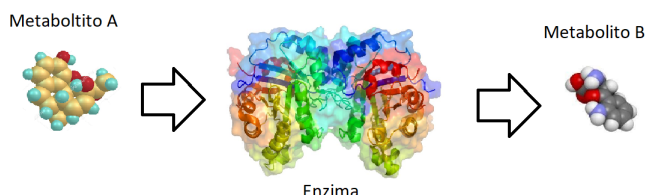


Figura 2-3: La enzima reacciona con el metabolito A (sustrato) proceso que implica una transformación del metabolito A en uno o varios metabolitos B (producto).

2.3.1. Vías Metabólicas

Las vías metabólicas son el nombre que se da a las sucesivas transformaciones sincronizadas que realizan un grupo de enzimas sobre los compuestos externos que entran en la célula, con la finalidad de obtener energía o estructuras químicas muy específicas. En las transformaciones intermedias a los compuestos se les denomina metabolitos, estos son trasferidos entre enzimas de una misma vía metabólica e incluso pueden ser transferida de una vía metabólica a otra, lo que hace en general del metabolismo celular un sistema interconectado y muy sofisticado de rutas. De manera muy general, las rutas metabólicas están compuestas de enzimas, metabolitos y reacciones químicas que se pueden agrupar en tres categorías:

- **Catabólicas:** Estas son vías metabólicas en las cuales las fuentes químicas externas, o nutrientes, que entran a la célula son transformadas mediante reacciones de degradación oxidativas, convirtiéndose en productos simples, que le permiten a la célula obtener energía.
- **Anabólicas:** Estas vías transforman moléculas precursoras, en otras mas grandes y complejas, como ácidos nucleicos y proteínas.
- **Anfibólicas:** Estas vías metabólicas son tanto catabólicas como anabólicas.

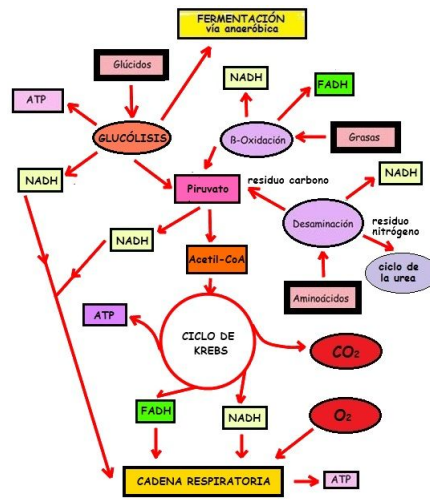


Figura 2-4: Representación de la interacción entre vías metabólicas que intervienen en la respiración celular. (Mónica Barboza, s.f) [13]

2.3.2. Redes Metabólicas

El estudio de las vías metabólicas se ha hecho durante décadas desde un enfoque reduccionista, debido a la modularidad que presenta el sistema metabólico celular. Los bioquímicos han logrado identificar con cierta precisión las vías metabólicas como sistemas de reacciones aisladas que cumplen tareas específicas. La gran cantidad de datos que se tiene sobre estas vías ha permitido estudiar el sistema metabólico desde la complejidad de sus interacciones, adaptándose el concepto de red metabólica, en la cual se agrupan todas las interacciones entre vías metabólicas [14]. estas redes pueden ser simples centradas en metabolitos o en enzimas, como se muestra en la Figura [2-5] a) y b), o bipartitas Figura [2-5] c).

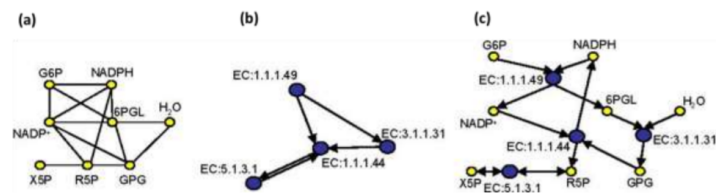


Figura 2-5: a) red centrada en metabolitos b) red centrada en enzimas c) red bipartita de enzimas y metabolitos [15]

En la Figura [2-6] se muestra un ejemplo de redes metabólicas basadas en metabolitos, en este caso, es una red de todo el metabolismo humano cuyos nodos están conectada si existe una reacción que permita convertir un metabolito en otro.

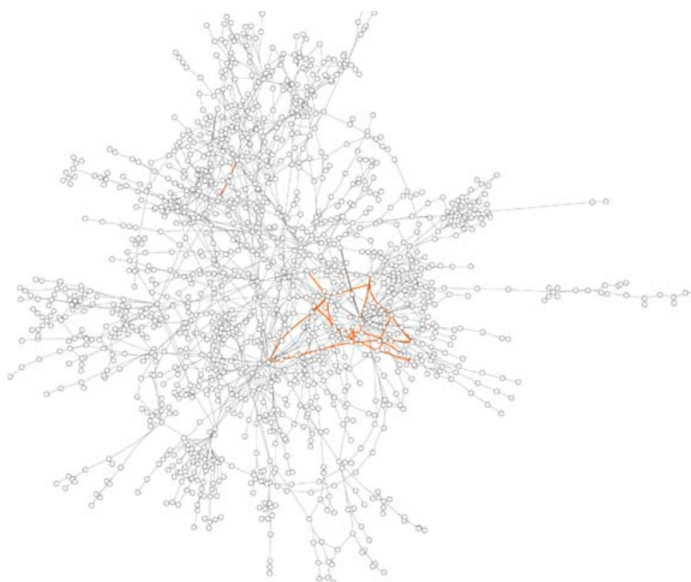


Figura 2-6: Red del metabolismo humano, basada en metabolitos, el subgrupo en naranja es el ciclo de Krebs, la primera ruta metabólica descubierta por Hans Krebs en 1937. [14]

Una de las características de las redes metabólicas es la existencia de metabolitos atractores como el ATP o el NADH, que están presentes en un número muy alto de reacciones[16] lo cual hace, que combinado con el hecho de que la mayoría de los metabolitos de la red están involucrado en un número bajo de reacciones, la red exhiba una distribución de conectividad de ley de potencias, es decir que la frecuencia $P(k)$ de los metabolitos que participan en k reacciones, es proporcional a $P(k) \sim k^{-\gamma}$, donde γ es una constante [17]. Ha sido un gran desafío para los biólogos comprender cómo han evolucionado las estructuras de estas redes, y explicar cuáles han sido las causas y las consecuencias de su diseño.

3 Metodología

3.1. Bases de Datos

A partir del árbol filogenético disponible en iTOL (Interactive Tree Of Life), se seleccionó una muestra de trece organismos, los cuales son representativos de linajes que tienen particular interés evolutivo Tabla[3-1], una aproximación de la relación evolutiva que existe entre estos organismos según iTOL se muestra en la Figura[3-1].

Tabla 3-1: Listado de organismos estudiados

Codigo Kegg	Nombre científico	Nombre comun	Reino	
hsa	Homo sapiens	Hombre	Eucariota - Mamífero	Pluricelular
ptr	Pan trogloditas	chimpancé	Eucariota - Mamífero	Pluricelular
mmu	Mus musculus	Raton domestico	Eucariota - Mamífero	Pluricelular
dme	Drosophila melanogaster	mosca de la fruta	Eucariota - Insecto	Pluricelular
cel	Caenorhabditis elegans	gusanos redondos	Eucariota - Nematodo	Pluricelular
ath	Arabidopsis thaliana		Eucariota - Planta	Pluricelular
lma	Leishmania major	Dolor oriental	Eucariota - Parasito	Unicelular
sce	Saccharomyces cerevisiae	levadura incipiente	Eucariota - Hongo	Unicelular
cac	Clostridium acetobutylicum	Organismo Weizmann	Bacteria	Unicelular
eco	Escherichia coli K-12 MG1655	Ecoli	Bacteria	Unicelular
sfx	Shigella flexneri 2457T	serotipo 2a	Bacteria	Unicelular
tac	Thermoplasma acidophilum		Arquea	Unicelular
sai	Sulfolobus acidocaldarius		Arquea	Unicelular

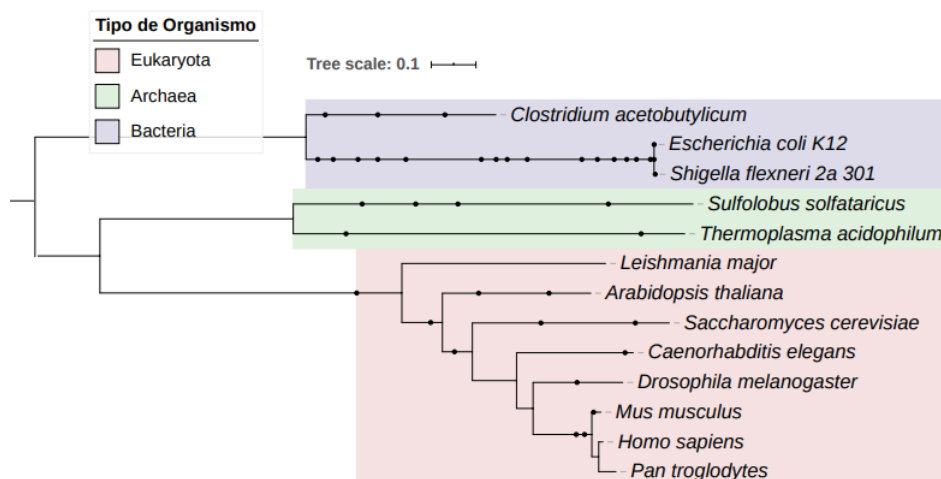


Figura 3-1: Árbol filogenético de los organismos estudiados, este diagrama fue obtenido de iTOL (Interactive Tree Of Life).

Los datos que se utilizaron para la construcción de las redes metabólicas fueron tomadas de KEGG (Kyoto Encyclopedia of Genes and Genomes) en la cual reposan unas robustas bases de datos de los principales vías metabólicas presentes en los organismos que estudiamos en este documento. En la Figura [3-2] se puede observar un diagrama de las vías metabólicas proporcionadas por KEGG, que pueden ser exportadas en formato XML y en los cuales se encuentra las relaciones enzima-enzima, bajo la etiqueta Ecnel, y enzima-compuesto.

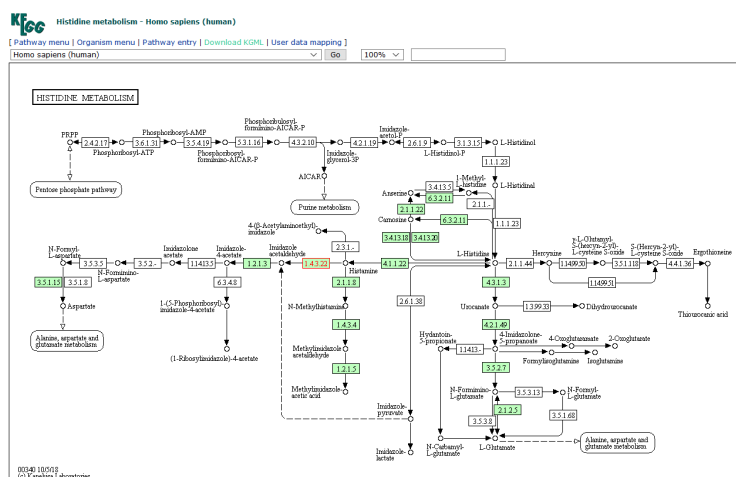


Figura 3-2: Diagrama del Metabolismo de la histidina del humano proporcionada por KEGG.

En este estudio reconstruimos la red metabólica centrada en enzimas de cada uno de los trece organismos de la Tabla[3-1], donde los nodos representan las enzimas del metabolismo, y

cuyos vínculos relacionan enzimas que catalizan pasos de reacción sucesivos. También se reconstruyen las redes bipartitas enzima-compuestos, en las cuales existen dos clase diferentes de nodos, las enzimas y los compuestos, y cuyos aristas relaciona los compuestos con las enzimas que lo catalizan.

La información proporcionada por KEGG, no contiene el numero EC(Enzyme Commission numbers) de las enzimas, si no la convención KO. Para obtener el numero EC es necesario hacer una búsqueda particular de cada enzima dentro de la base de datos de KEGG.

3.2. Teoría de grafos

Los sistemas complejos, vistos desde la teoría de grafos, poseen propiedades estructurales o topológicas que nos permiten caracterizarlos y analizarlos, estas dan cuenta de como se conectan sus elementos (nodos) unos con otros. Algunas de las medidas topológicas mas importantes son:

3.2.1. Medidas de grafo

- **El grado nodal y la conectividad:**

Una de las medidas elementales que se puede tomar de un nodo es el número de vecinos que tiene, es decir el número de nodos con los que esta conectado, este numero seria la conectividad o el grado del nodo. En las redes dirigidas podemos definir dos tipos diferentes de grado, el grado de entrada que representaremos por k_{in}^i y que es el número de vecinos cuyas aristas vienen en dirección al nodo n_i y el grado de salida representado por k_{out}^i que es el número de vecinos cuyas aristas van en dirección opuesta al nodo. Estos dos tipos de grado pueden tener o no comportamientos distintos cuya interpretación es caso de estudio de cada sistema en particular.

- **Grado Promedio:**

El grado como lo mencionamos anteriormente es una medida elemental, y muy potente dentro del análisis de las redes, razón por la cual existen varias mediciones asociadas con esta, entre ellas el grado promedio y la distribución de grado, que usualmente son de mucho interés en el estudio de sistemas complejos. El grado promedio \bar{k} es una medida característica de la red en general (“medida macroscópica”) que se puede calcular para redes no dirigidas como:

$$\bar{k} = \frac{\sum_{i=1}^g k^i}{g} = \frac{2L}{g} \quad (3-1)$$

Donde k^i es el número de vínculos que tiene el nodo n_i , g es el número total de nodos de la red, y L el número total de vínculos.

Para el caso del grafo dirigido, el vínculo existente entre dos nodos se cuenta solo una vez, con esto la expresión cambia como:

$$\bar{k} = \frac{L}{g} \quad (3-2)$$

■ **Distribución de grado:**

La distribución de grado es tal vez una de las propiedades mas importantes que caracterizan la estructura de una red. Siendo esta una distribución de la probabilidad de encontrar un nodo n_i con un valor de grado $P(k^i)$ en la red. Esta probabilidad se obtiene del conteo de nodos n_i con grado k^i dividido por el número total de nodos de la red g . En una red dirigida es claro que tenemos dos distribuciones de grado distintas, las correspondientes a los grado de entrada y los grados de salida de los nodos.

■ **Densidad:**

Otra medida importante de una red es su densidad, que se define como el número de vínculos o aristas totales L sobre el número de vínculos posibles L_p . Para calcular el número de vínculos posibles L_p de una red no dirigida, que es el número de aristas que debería tener una red con g nodos para ser completo, se utiliza la expresión:

$$L_p = \frac{g(g-1)}{2} \quad (3-3)$$

Y para el caso de una red dirigida el número de vínculos posibles se puede calcular como :

$$L_p = g(g-1) \quad (3-4)$$

Con esto entonces la densidad será calculada como:

$$\rho = \frac{L}{L_p} \quad (3-5)$$

■ **Coefficiente de Asortatividad**

Este coeficiente esta asociado con el coeficiente de correlación de Pearson, el cual cuantifica la preferencia que tienen los nodos de la red a establecer vínculos con nodos que le son similares en alguna característica, como el grado. Este coeficiente se define como:

$$r = \sum_{jk} \frac{jk(e_{jk} - q_j q_k)}{P^2} \quad (3-6)$$

Con

$$P^2 = \sum_k K^2 q_k - \left[\sum_k k q_k \right]^2 \quad (3-7)$$

Este coeficiente puede tener valores entre -1 y 1. Cuando la característica es el grado de los nodos, un Coeficiente de Asortatividad positivo nos indica que los nodos de la red tiene preferencia en establecer vínculos con nodos de grado similar, en el caso de que $r = 0$ la red no es asortativa, pero cuando la red tiene un coeficiente negativos la red es Disortativa y nos sugiere que los nodos de la red de grado alto o atractores tienen preferencia por vincularse con nodos de bajo grado y ellos a su vez a vincularse con nodos de grado alto, como consecuencia esta red muestra un carácter de concentrador haciéndola desordenada.

■ Reciprocidad

La reciprocidad R de una red es una medida de la probabilidad de que dos nodos tenga una interacción reciproca, es decir que estén conectadas por medio de aristas dirigidas de forma mutua, tanto del nodo n_a al n_b como del nodo n_b al n_a . Esta medida de probabilidad se puede calcular como el cociente entre el número de aristas reciprocas L^{\leftrightarrow} y el número total de aristas L de la red:

$$R = \frac{L^{\leftrightarrow}}{L} \quad (3-8)$$

■ Coeficiente de Clusterización promedio

El coeficiente de clusterizacion o de agrupamiento de un nodo nos da información del nivel de interconexión que este tiene con sus vecinos y la red en general. Para cuantificar la agrupación de un nodo n_i se calcula la fracción de triángulos $\text{Tr}(n_i)$ que existen a través del nodo n_i como:

$$C_i = \frac{2\text{Tr}(n_i)}{k^i(k^i - 1)} \quad (3-9)$$

Donde k^i es el grado del nodo n_i . Ahora el Coeficiente de Clusterizacion promedio se calcula como el valor promedio de este coeficiente para todos los nodos de la red, es decir:

$$\hat{C} = \frac{\sum_{i=1}^g C_i}{g} \quad (3-10)$$

■ Transitividad

La transitividad es una propiedad de las redes que considera las relaciones entre tríos de nodos. Por definición una relación es transitiva si para un grupo de nodos n_i , n_j y n_k siempre que exista una interacción entre $n_i \rightarrow n_j$ y $n_j \rightarrow n_k$, entonces se da que $n_i \rightarrow n_k$ Figura[3-3]. En este estudio utilizamos la función *transitivity* de la librería *NetworkX* en el cual se calcula la transitividad como la fracción de triángulos transitivos sobre el número total de triángulos posibles presentes en la red.

$$T = 3 \times \frac{\# \text{Triángulos Transitivos}}{\# \text{Tríadas}} \quad (3-11)$$

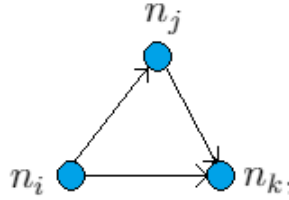


Figura 3-3: Diagrama de una relación transitividad entre tres nodos.

En un sistema complejo sus elementos pueden o no tener la misma importancia, en algunas circunstancias remover un elemento de un sistema puede alterara de forma significativa su funcionamiento. Al trabajar el sistema como una red podemos cuantificar el nivel de importancia de un nodo por medio de su centralidad. En la actualidad existe un número grande de centralidades que podemos medir de un nodo, alguna de ellas son:

■ Centralidad de grado

El grado de un nodo nos da información que usualmente es relevante para entender la importancia de un nodo en el sistema, debido a que un nodo muy conectado puede ser muy influyente en el sistema, es por eso que una medida de centralidad es el grado del nodo K_i

$$C_d(n_i) = K_i \quad (3-12)$$

Para hacer esta medida comparable para redes de tamaños diferentes se hace una estandarización de la medida:

$$C'_d(n_i) = \frac{K_i}{g-1} \quad (3-13)$$

Aquí C'_d es independiente de g .[?]

■ Centralidad de cercanía

Esta centralidad esta basada en la cercanía de un nodo a los demás nodos de la red, en este sentido un actor es central si interactúa rápidamente con todos los nodos. Una forma de calcular esta centralidad es:

$$C_c(n_i) = \left[\sum_{j=1}^g d(n_i, n_j) \right]^{-1} \quad (3-14)$$

Esta formulación es propuesta por Sabidussi (1966) donde $d(n_i, n_j)$ es el número de líneas geodésicas que unen los nodos i y j , entonces $\sum_{j=1}^g d(n_i, n_j)$ sera la distancia total a la que esta n_i de los demás nodos.

Dado a que la suma de distancias depende del número de nodos de la red a los que el nodo n_i tiene acceso, es decir $g-1$ se hace la estandarización de esta medida como:

$$C'_c(n_i) = \frac{g-1}{\left[\sum_{j=1}^g d(n_i, n_j) \right]} \quad (3-15)$$

■ Centralidad de intermediación

Es un centralidad basada en la importancia que tiene un nodo respecto a su capacidad de conectar otros nodos a la red, su medida cuantifica el número de veces que dicho nodo actúa como puente entre otro par de nodos del sistema. Esta centralidad se calcula como una suma de probabilidad definida como:

$$C_B(n_i) = \sum_{j < K} \frac{g_{ik}(n_i)}{g_{ik}} \quad (3-16)$$

Donde $g_{jk}(n_i)$ es el número de geodésicas que unen a dos actores n_j y n_k , y que contiene el nodo n_i , dado que la probabilidad de escoger cualquier geodésica es la misma, se divide entonces por el número total de geodésicas que unen a los nodos n_j y n_k .

Derivado de estas medidas de centralidad de nodo también podemos calcular centralidades de grupo de la red, las cuales proporcionan valores característico generales de la red basada en la centralidad de sus nodos, entre estas medidas encontramos:

■ Cercanía de grupo

Esta centralidad de grupo esta relacionada con la centralidad de cercanía de los nodos, y se puede calcula utilizando el indice de Freeman(1979) :

$$C_c = \frac{\sum_{i=1}^g [C'_c(n^*) - C'_c(n_i)]}{[(g-2)(g-1)]/(2g-3)} \quad (3-17)$$

Donde $C'_c(n^*)$ es la centralidad de cercanía estandarizada de mayor valor dentro de la red.

■ Intermediación de grupo

Al igual que la centralidad de cercanía de grupo, la centralidad de intermediación de grupo esta relacionada con la intermediación de los nodos dentro de la red, siendo este una reflejo de su heterogeneidad. Una forma de calcular este intermediación es mediante el uso de la siguiente ecuación:

$$C_B = \frac{2 \sum_{i=1}^g [C'_B(n^*) - C'_B(n_i)]}{(g-1)} \quad (3-18)$$

Siendo $C'_B(n^*)$ la centralidad de intermediación de mayor valor dentro de la red.

3.3. Propiedades estadísticas de las redes complejas

Una red puede ser una estructura muy compleja y difícil de analizar, para poder tener una visión simplificada y comprensible se toman algunas medidas estadísticas que permitan capturar su información estructural, algunas de estas medidas son:

3.3.1. Distribuciones de grado

Una de las propiedades mas interesantes que se pueden obtener del estudio de las redes es su distribución de probabilidad de grado $P(k)$, en la cual se expresa las probabilidades que se tiene de encontrar al azar un nodo n_i con un grado k dado. La distribución de probabilidad suele ser importante por que reúne mucha información sobre la estructura de las redes, por ejemplo en una red simple usualmente los nodos tienen grado similar, mas sin embargo en muchas redes de sistemas reales la distribución de grado nos suele indicar que la mayoría de los nodos tiene un grado bajo, y existe solo un pequeño grupo de nodos que tiene grados altos. Estas distribuciones usualmente están asociadas a características estructurales o topologías diferentes como:

- **Topología libre de escala:** En estas las distribuciones de grado describe una ley de potencias:

$$P(k) = Ae^{-\gamma} \quad (3-19)$$

Muchas de los sistemas de redes en la naturaleza presentan esta topología, cuyas características principales es que son usualmente robustas, resilientes y eficientes en la transmisión de información [18].

- **Topología de Poisson:** Usualmente son redes de tipo Erdős-Renyi, las cuales son grafos aleatorios, en los cuales un nuevo nodo se enlaza con igual probabilidad con cualquier otro nodo de la red, este tipo de redes sigue una distribución de probabilidad tipo Poisson:

$$P(k) = e^{-z} \frac{z^k}{k!} \quad (3-20)$$

En la naturaleza son limitados los sistemas que tienen un distribución de este tipo [19].

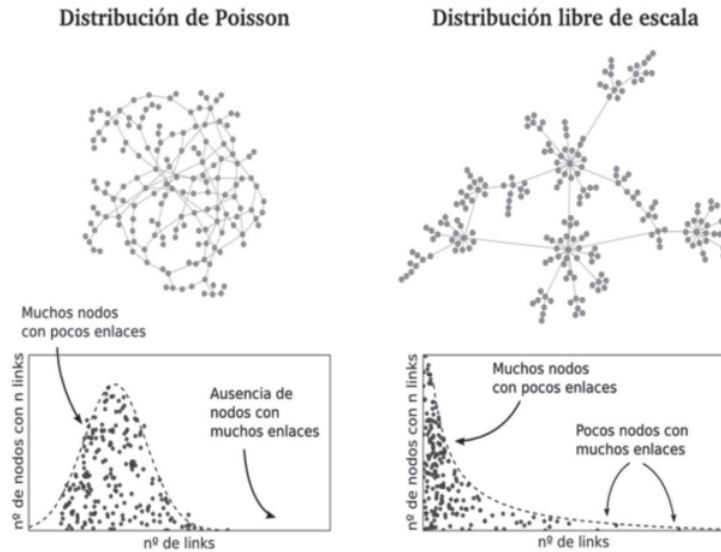


Figura 3-4: Ejemplos de como se ven visualmente y como es el diagramas de las distribución de grado de Poisson y libre de escala.(Javier Toret. 2015)[18]

- **Topología Exponencial:** Este tipo de redes tiene una distribución de probabilidad exponencial:

$$P(k) = Ae^{-\alpha k} \quad (3-21)$$

En la naturaleza usualmente esta topología esta relacionada con sistemas que evolucionan en el tiempo, en donde cada nuevo nodo tiene igual probabilidad de vincularse con cualquier otro nodo en la red [20].

3.3.2. Entropía

En una distribución de probabilidad asociada a un conjunto de N número de eventos mutuamente excluyentes, cuyas probabilidades son $\{p_1, p_2, \dots, p_N\}$, se puede definir la *sorpres*a de que ocurra un evento dado en un ensayo como:

$$I(p_n) = -k \ln p_n \quad (3-22)$$

Donde k es la constante de Boltzmann, $k = 1,38066 \times 10^{23} J/K$. Esta cantidad se relaciona con la falta de información que se tiene sobre la ocurrencia del evento previo a un ensayo y su compartimento indica que entre más probable sea un evento menos información contiene [21].

La *sorpres*a promedio o el valor medio de la información del conjunto de eventos se calcula por medio de la fórmula de la entropía de Shannon o entropía estadística que se define como:

$$H = -k \sum_{n=1}^N p_n \ln p_n \quad (3-23)$$

Dado a que todos los eventos o estados son accesibles la suma de las probabilidades debe ser uno:

$$\sum_{n=1}^N p_n = 1 \quad (3-24)$$

La entropía puede considerarse una medida de dispersión, en tanto que es máxima cuando las distribuciones de probabilidades son uniformes en el caso discreto o se aproxima a una distribución gaussiana en el caso continuo, y es mínima cuando la distribución es una delta. Las situaciones para el caso discreto se muestran en la figura 3-5 en donde en A) la entropía es máxima y en B) la entropía es cero, por convención se asume que $\lim_{p \rightarrow 0} p \ln p = 0$.

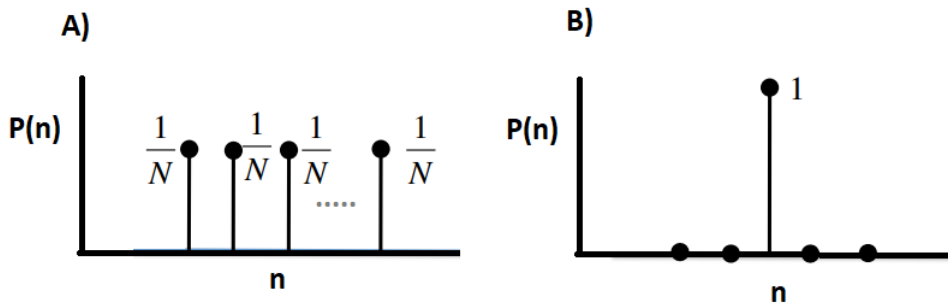


Figura 3-5: A) Distribución de probabilidad discreta con entropía máxima. B) Distribución discreta con entropía nula.

La entropía en comparación con la varianza, que es otra medida de dispersión, no se ve afectada por la posición o el orden de las deltas, por ejemplo en la figura [3-6] se ilustran

dos situaciones en las cuales las distribuciones aunque son diferentes la probabilidad de cada estado se conserva, por lo tanto las entropías son iguales pero la varianza es diferente.

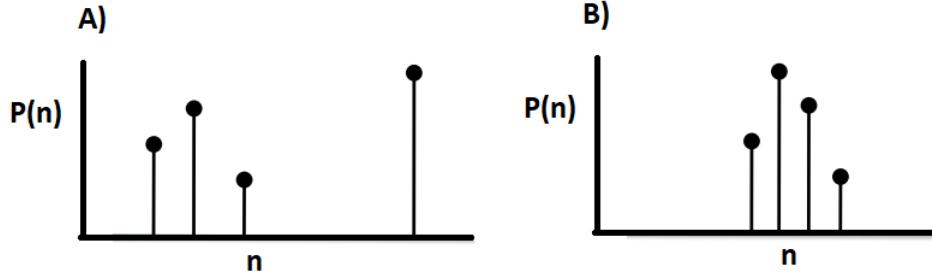


Figura 3-6: Distribuciones de probabilidad discreta con entropías iguales y varianza diferente.

En los sistemas complejos la constante de Boltzmann $k = 1$ por que no aparece explícitamente en la formula de la entropía.

$$S = - \sum_{n=1}^N p_n \ln p_n \quad (3-25)$$

3.3.3. Divergencia de Kullback-Leibler

La divergencia de Kullback-Leibler (D_{KL}) usualmente se considera como una métrica o distancia relacionada con la similitud entre dos distribuciones de probabilidad $P = \{p_1, p_2, \dots, p_i\}$ y $Q = \{q_1, q_2, \dots, q_i\}$, y se calcula como el promedio ponderado de la diferencia logarítmica de las dos distribuciones, ecuación (3-26).

$$D_{KL} = \sum_i p_i \ln \left(\frac{p_i}{q_i} \right) \quad (3-26)$$

Esta divergencia solo se puede calcular si se cumple las siguientes condiciones:

- $\sum_i p_i = 1$ y $\sum_i q_i = 1$
- $q_i > 0$ para cualquier i

Si se obtiene $0 \ln 0$ este valor se toma como cero [22].

3.4. Análisis comparativo de Redes

Pese a que el modelado y el análisis de redes durante estos últimos años ha tenido gran éxito al estudiar sistemas interconectados, aun existe una dificultad metodológica respecto

al análisis comparativo que se puede hacer entre redes, es por ello que, al momento en el cual fue presentado este trabajo, no existe una metodología única para llevar a cabo este tipo de análisis. Mas sin embargo, se han hecho esfuerzos para poder llevar a cabo este tipo de análisis en algunos sistemas de importancia practica en campos de la ciencia, como la química, el estudio de las redes sociales y la neurociencia [23].

En la sección 3.2.1 se muestran algunas medidas topológicas de las redes, que en un principio se pueden utilizar para comparar y cuantificar la similitud o disimilitud entre un grupo de redes con respecto a un conjunto de medidas dadas. En general, una comparación entre redes, dada una medida estructural en especifico no es fiable, debido a que no se garantiza que su similitud sea producto de los mismos mecanismos estructurales. Es por ello que en este estudio se utiliza el Análisis de Componentes Principales, o ACP, que es una técnicas de la estadística la cual permite analizar un conjunto de variables estructurales de forma simultanea.

3.4.1. Análisis de Componente Principales - ACP

El ACP es una metodología algebraico/estadística que permite la reducción de la dimensionalidad de un conjunto de datos, y consiste en la obtención de nuevas variables o componentes, en las cuales este contenida la mayor parte de la información o variabilidad de los datos[24]. El procedimiento por el cual se obtienen estos componentes es mediante la matriz de covarianzas o matriz de correlaciones, en la cual se aplica una transformación lineal que da como resultado un nuevo conjunto de variables, cuyo objetivo es:

- Que sean linealmente independientes.
- Tenga un porcentaje de variabilidad explicada conocida y diferenciada entre ellas.
- Y que se conserven la variabilidad total de los datos.

3.4.2. Asimetría de una distribución.

Otra medida estadística implementada en este estudio que permite hacer una comparativa de las distribuciones de probabilidad de grado en cuanto a su forma, es el grado de asimetría que la distribución presenta al rededor de la media, esta se mide con el coeficiente de asimetría de Fisher, el cual parte del uso del tercer momento, ecuación (3-27), este coeficiente permite mantener el signo de las desviaciones con respecto a la media, información que es relevante en este estudio.

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{3/2}} \quad (3-27)$$

Donde σ es la desviación estándar y μ_3 es el tercer momento al rededor de la media.

3.4.3. Correlación de Pearson

El coeficiente de correlación de Pearson mide el grado de relación que hay entre dos variables aleatorias por medio de la dependencia lineal, este se define como:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (3-28)$$

Donde σ_X y σ_Y son las desviaciones estándar de las variables X y Y respectivamente, y μ_X y μ_Y son las medias.

4 Resultados y discusión

En este capítulo se hace un análisis comparativo de las estructuras de las redes metabólicas de un grupo de 13 organismos, a partir de sus características de conectividad, con medidas de propiedades topológicas microscópicas.

4.1. Análisis comparativo de redes metabólicas implementando un ACP

Una forma en la que podemos tener un panorama general de las diferencias o similitudes entre los sistemas metabólicos celulares de los organismos estudiados, es mediante la comparación estructural de sus redes metabólicas. En la Tabla[4-1] tenemos algunas de las medidas “Macroscópicas” que tomamos de cada una de las redes.

Tabla 4-1: Principales propiedades estructurales medidas de las redes metabólicas de los organismos estudiados

Nombre	N nodos	N arcos	Grado promedio	Densidad	Reciprocidad	Asortatividad	Intermediación grupo	Cercanía grupo	Transitividad	Cluster promedio
hsa	803	3066	3,82	0,0095	0,27	0,24	0,080	0,159	0,275	0,427
ptr	811	3059	3,77	0,0093	0,26	0,24	0,083	0,159	0,278	0,432
mmu	796	2953	3,71	0,0093	0,26	0,21	0,084	0,157	0,275	0,424
cel	483	1373	2,84	0,0118	0,27	0,27	0,067	0,163	0,262	0,351
dme	522	1437	2,75	0,0106	0,27	0,24	0,090	0,145	0,240	0,362
eco	682	1811	2,66	0,0078	0,19	0,31	0,084	0,128	0,250	0,344
ath	787	2069	2,63	0,0067	0,25	0,26	0,078	0,103	0,263	0,393
sfx	610	1547	2,54	0,0083	0,19	0,27	0,067	0,137	0,244	0,323
sce	495	1253	2,53	0,0102	0,25	0,29	0,074	0,128	0,265	0,337
cac	414	877	2,12	0,0103	0,22	0,36	0,059	0,116	0,237	0,287
lma	288	592	2,06	0,0143	0,25	0,33	0,085	0,146	0,234	0,281
tac	240	471	1,96	0,0164	0,28	0,37	0,043	0,140	0,303	0,302
sai	276	485	1,76	0,0128	0,25	0,33	0,029	0,154	0,213	0,235

La Tabla[4-1] se encuentra ordenada de mayor a menor de acuerdo al grado promedio, esta medida llama especialmente la atención debido a la jerarquización que sugiere de los organismos, donde encontramos los taxones evolutivamente cercanos, en posiciones próximos dentro de la tabla. El grado promedio resulta ser una de las mejores de estas medidas estructurales que clasifica organismos aproximándonos a la posición dentro sus rutas evolutivas.

Hay un conjunto de variables adicionales con las que podemos complementar la tabla comparativa (Tabla[4-1]) aprovechando información mas detallada que podemos extraer de los sistemas, y que de forma intuitiva pueden sustentarse bajo algunos aspectos evolutivos. Un subconjunto de estas variables adicionales son las centralidades de un grupo de enzimas presentes en todos los organismos estudiados, esta medición nos permite visualizar los cambios en la importancia de estas enzimas dentro de la redes producto de las presiones evolutivas Tabla[4-2].

Tabla 4-2: Medidas de centralidad de algunas enzimas presentes en las Redes metabólicas.

Nombre	Centralidad de grado in			Centralidad de Intermediacion			Centralidad de grado out			Centralidad de cercania		
	[EC:2.1.2.1]	[EC:2.6.1.16]	[EC:4.3.2.2]	[EC:2.2.1.1]	[EC:4.3.2.2]	[EC:2.5.1.16]	[EC:2.1.2.1]	[EC:2.6.1.16]	[EC:4.3.2.2]	[EC:4.2.1.11]	[EC:2.2.1.1]	[EC:5.3.1.1]
sai	3	2	4	0,00087	0,0012	0,00016	4	2	1	0,098	0,022	0,020
tac	7	2	4	0,00109	0,0277	0,00019	7	4	1	0,092	0,026	0,020
eco	7	3	7	0,00992	0,0064	0,01734	13	7	4	0,093	0,029	0,023
hsa	11	5	9	0,00726	0,0047	0,00643	19	5	6	0,113	0,089	0,083
ptr	11	5	9	0,00697	0,0050	0,00618	19	5	6	0,112	0,087	0,081
ath	9	4	5	0,04534	0,0075	0,00257	12	6	5	0,072	0,061	0,054
mmu	11	5	9	0,00892	0,0059	0,00628	20	5	6	0,113	0,111	0,086
cac	6	3	5	0,00353	0,0087	0,00087	8	6	2	0,081	0,037	0,028
cel	11	5	5	0,00955	0,0126	0,00126	13	6	5	0,104	0,064	0,058
dme	11	5	7	0,00642	0,0138	0,00056	15	6	5	0,103	0,063	0,060
lma	4	4	4	0,01183	0,0059	0,01589	8	4	3	0,091	0,035	0,026
sce	11	4	6	0,00819	0,0134	0,00269	11	5	4	0,081	0,043	0,040
sfx	7	3	7	0,00378	0,0062	0,01753	13	7	4	0,091	0,029	0,022

Otras variables consideradas en el estudio están relacionadas con el porcentaje de nodos de determinada clase dentro de la red, como lo son nodos Receptores, Transmisores y Ordinarios [2.2.1]. Esta clasificación, vista desde el sistema metabólico, distingue una buena parte de las enzimas que se encuentran al principio (Enzimas Transmisora) o final (Enzima Receptora) de un proceso de transformación de un alimento. En la tabla [4-3] se muestran el porcentaje de estos tipos de enzima en la red metabólica.

Por otra parte la diferencia de grado es una medida que resulta interesante en este sistema, de acuerdo a esta podemos establecer tres tipos de enzimas:

- Enzima de Fusión: Si el nodo n_i cumple que $k_{in}^i > k_{out}^i$
- Enzima de Bifurcación: Si el nodo n_i cumple que $k_{in}^i < k_{out}^i$
- Enzima de Paso: Si el nodo n_i cumple que $k_{in}^i = k_{out}^i$

Aunque algunas de estas medidas por si solas permitan una clasificación relacionada con las distancias evolutivas de estos organismos, y suponemos que toman algún sentido dentro de los mecanismos evolutivos de estas redes, es necesario tener en cuenta que dichos mecanismos pueden llegar a ser muy diferentes para caminos evolutivos distintos y puede verse reflejado en un conjunto de características estructurales en especifico para cada trayectoria evolutiva,

Tabla 4-3: Porcentaje de enzimas pertenecientes a las clases; Receptor($k_{in}^i > 0$ y $k_{out}^i = 0$), Transmisor($k_{in}^i = 0$ y $k_{out}^i > 0$) y Ordinaria($k_{in}^i > 0$ y $k_{out}^i > 0$), Fusión $k_{in}^i > k_{out}^i$, Bifurcación $k_{in}^i < k_{out}^i$ y Paso $k_{in}^i = k_{out}^i$.

Nombre	%Paso	%Fusion	%Bifurcacion	%receptor	%transmisor	%ordinario
sai	29,3	35,1	35,5	15,9	15,6	68,5
tac	29,2	35,4	35,4	17,9	15,8	66,3
eco	25,1	37,7	37,2	12,3	12,3	75,4
hsa	17,8	40,6	41,6	9,8	9,0	81,2
ptr	17,9	40,6	41,6	9,9	9,4	80,8
ath	23,3	39,3	37,5	13,1	12,3	74,6
mmu	18,3	40,1	41,6	9,4	9,0	81,5
cac	28,3	35,5	36,2	15,9	13,8	70,3
cel	19,9	39,1	41,0	12,8	13,7	73,5
dme	20,9	40,0	39,1	12,8	12,5	74,7
lma	22,6	39,9	37,5	17,7	18,8	63,5
sce	24,2	37,2	38,6	12,9	11,7	75,4
sfx	23,9	37,9	38,2	13,6	13,1	73,3

lo que puede ocasionar que al evaluar separadamente parámetros estructurales como, por ejemplo, el número de nodos (**N nodos**) aparezcan mas próximos o relacionados organismos con trayectorias que a priori no tiene tal nivel de proximidad; mas sin embargo, sabemos que esta propiedad es importante desde el punto de vista evolutivo debido al hecho de que los sistemas metabólicos, entre otros aspectos, tienden a agregar mas enzimas a su estructura. Esto en términos prácticos sugiere una búsqueda mas profunda en las variables que nos permita determinar si un conjunto de propiedades estructurales esta mas relacionada con una trayectoria evolutiva en particular.

Usando el coeficiente de correlación de Pearson, en la Figura[4-1] observamos que existen propiedades o variables estructurales fuertemente correlacionadas, dado esto, reducimos la dimensionalidad teniendo como criterio una correlación positiva superior al 90 %. Con esto tenemos:

- Grado promedio \approx número de arcos en un 97 % $\approx k_{out}$ de la enzima [EC:2.1.2.1] en un 97 % \approx Cluster promedio en un 94 % $\approx k_{in}$ de la enzima [EC:4.3.2.2] en un 92 % $\approx K_{out}$ de la enzima [EC:4.3.2.2] en un 91 % \approx %Bifurcacion en un 93 % \approx %ordinario en un 92 % \approx Cercania [EC:2.2.1.1] en un 91 % \approx Cercania [EC:5.3.1.1] 91 %
- Cercanía de grupo \approx Cercanía de la enzima [EC:4.2.1.11] 91 %
- %fusion $\approx k_{in}$ de la [EC:2.6.1.16] en un 90 %

El diagrama de la matriz de correlaciones de Pearson del conjunto de variables reducidas se muestra en la Figura[4-2]. Con este nuevo conjunto de variables y aplicando un ACP (Método de Componentes Principales) Figura[4-3] podemos hacer una comparativa general de las estructuras metabólicas de estos organismos.

4.1 Análisis comparativo de redes metabólicas implementando un ACP 27

Del ACP obtenemos una reducción de nuestro conjunto de diecisiete variables estructurales a dos nuevas variables no correlacionadas, con un porcentaje superior al 80 % de variable explicada, Figura[4-3]. Con esta síntesis de la información estructural de las redes metabólicas, podemos notar la proximidad que exhiben los sistemas metabólicos de organismos con distancias evolutivas cercanas. Calculando la Distancia euclidiana entre organismos, Tabla[4-4], podemos cuantificar la similitud o proximidad entre las estructuras de las redes metabólicas respecto a las variables medidas.

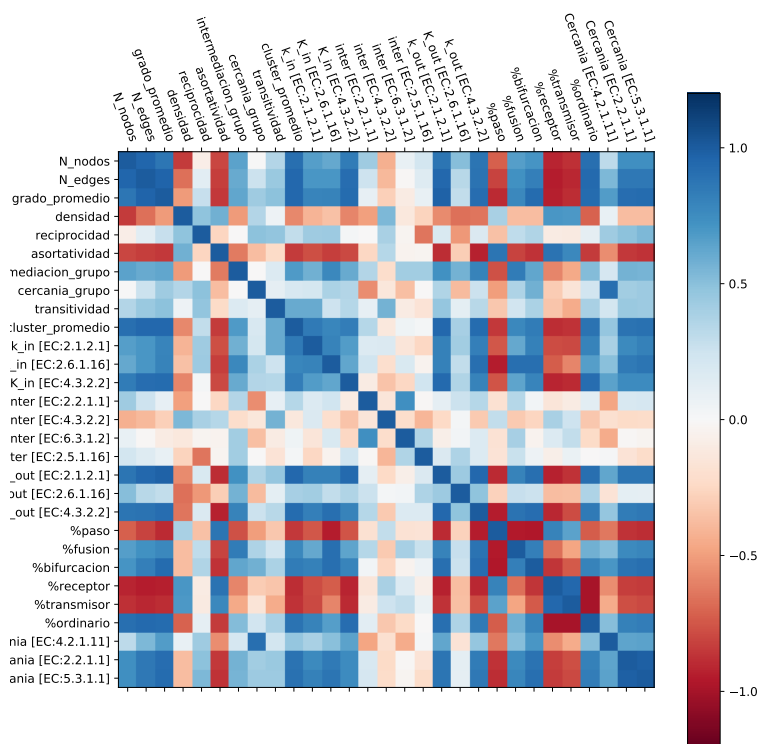


Figura 4-1: Diagrama de correlación que indica la correlación de Pearson entre las variables seleccionadas como más informativas para la caracterización de las redes.

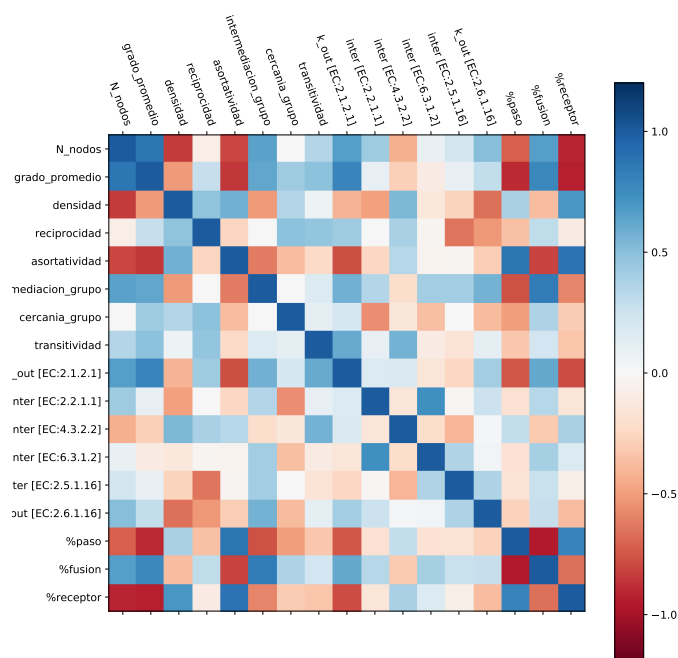


Figura 4-2: Diagrama que indica la correlación de Pearson entra el conjunto de variables estructurales cuya correlación es inferior al 90 %

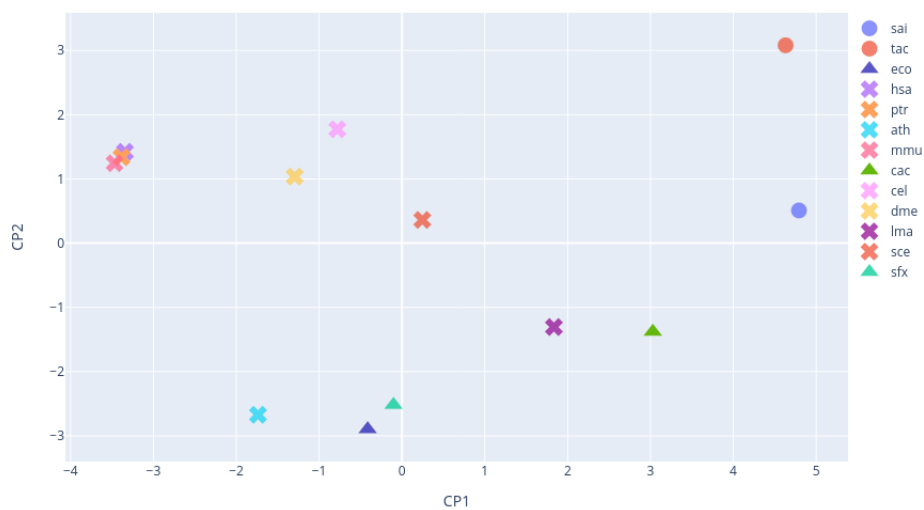


Figura 4-3: ACP aplicado sobre el conjunto de variables estructurales reducidas bajo el criterio de correlación inferior al 90 %. ● Arqueas, ▲ Bacterias y x Eucaritas.

Tabla 4-4: Distancia euclidiana relativa al taxón Pan Troglodyte (ptr), dentro del sistema de coordenadas de las componentes principales seleccionadas del ACP. En gris los taxones del dominio eucariota.

Nombre	Distancia euclidiana
ptr	0,00
hsa	0,09
mmu	0,13
dme	2,11
cel	2,64
sce	3,76
ath	4,34
sfx	5,06
eco	5,17
lma	5,85
cac	6,97
tac	8,20
sai	8,22

4.1.1. Discusión

Del cálculo de las distancias euclidianas relativas al taxón Pan Troglodyte (ptr) en el ACP, donde se resumen todas las variables topológicas estudiadas, es posible comparar la similitud estructural entre las redes metabólicas en un sentido muy amplio.

De este análisis se puede observar dos hechos significativos, el primero es que organismos diferentes tienen sistemas metabólicos diferentes, y el segundo, resultado de hacer una comparación con las distancias evolutivas en el árbol filogenético y la distancia euclidiana únicamente de los organismos Eucariotas, se puede ver que estas distancias se corresponden, de este hecho se puede concluir que las redes metabólicas de los organismos exhiben similitudes y diferencias topológicas que corresponden con procesos evolutivos diferentes, equivalentes a diferencias fenotípicas, genotípicas y arqueológicas producidas por presiones evolutivas.

Otro aspecto relevante que surge de este análisis, el cual queda por estudiar a profundidad en un estudio futuro, es la separación angular que exhiben los organismos en la distribución del ACP, Figura [4-3], la cual cualitativamente sugiere que las redes metabólicas no evolucionan preferentemente hacia alguna característica topológica en particular, y que de hecho esta diferencia angular aparente es mayor cuanto mas distancia evolutiva les corresponda a dos

organismos, lo que indicaría que la evolución de las redes metabólicas es estructuralmente diversa.

4.2. Análisis comparativo del grado nodal de las Redes metabólicas

El grado nodal, como lo mencionamos en el capítulo anterior, es una de las medidas estructurales de mayor importancia en la caracterización y análisis de redes. En el estudio de las redes metabólicas, el grado nodal es de particular interés debido a la dificultad que plantea la generación y consolidación de un nuevo vínculo, y lo que implica en la evolución de los organismos. El grado nodal de una enzima en éste sistema nos proporciona información sobre el número de enzimas con las que esta comparte compuestos, y en general sobre los procesos de transformación coordinada que se lleva acabo entre pares de enzimas, es así que el grado nodal y en general la conectividad de la red nos habla de procesos de sofisticación en el metabolismo.

4.2.1. Las tuplas $[k_{out}, k_{in}]$

Para comenzar a indagar el papel del grado nodal enzimático en los procesos evolutivos del metabolismo, distinguimos que dentro de este sistema una enzima puede tener dos tipos diferentes de grado k_{in} y k_{out} . Una forma inicial de agrupar esta información de grado es mediante la definición de las tuplas $[k_{out}, k_{in}]$. En esta sección se hace un análisis de estas tuplas tanto para las redes centradas en enzimas como para las redes bipartitas.

Redes Centradas en Enzimas

En este tipo de redes metabólicas el grado de entrada k_{in} de una enzima cuantifica el número de enzimas que envían compuestos hacia ella y el grado de salida k_{out} el número de enzimas a las cuales ella les envía compuestos ya catalizados. En este estudio es de mucho interés analizar el comportamiento de la tupla $[1, 1]$ el cual expresa la situación en la cual una enzima recibe un único producto de otra enzima y envía un único sustrato a otra enzima.

En la figura 4-5 se muestra un histograma comparativo en el cual se expresa el porcentaje de enzimas con tupla $[1, 1]$ dentro de la red metabólica de los diferentes taxones, siendo esta tupla la de mayor presencia en todos los organismos. En esta comparativa podemos observar una diferencia notable entre organismos eucariotas y procariotas, donde estos últimos presentan una prevalencia mayor de enzimas con esta tupla.

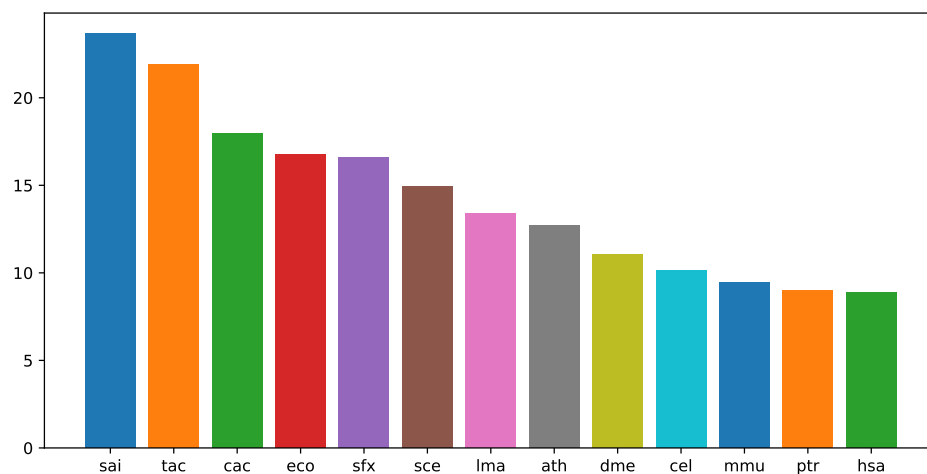


Figura 4-4: Histograma que indica el porcentaje de nodos con tupla $[k_{out}, k_{in}] = [1, 1]$ presentes en las redes metabólicas de los organismos estudiados.

Tabla 4-5: Tabla de entropías asociadas con la dispersión de las tuplas $[k_{out}, k_{in}]$ en las redes metabólicas centradas en enzimas de los organismos estudiados. En gris los taxones del dominio eucariota.

Org	Entropía
hsa	6,25
ptr	6,22
mmu	6,17
cel	5,58
ath	5,50
dme	5,44
eco	5,39
sce	5,34
sfx	5,32
cac	4,88
lma	4,80
tac	4,50
sai	4,27

Para cada una de las redes metabólicas estudiadas, existe una distribución particular de las tuplas $[k_{out}, k_{in}]$ que refleja en general diferencias en los procesos que se llevan a cabo en el metabolismo. Es posible comparar, entre los diferentes taxones, la distribución de esta me-

dida estructural aplicando la entropía como una medida de dispersión de las tuplas dentro de las redes metabólicas. Esta medida de entropía está calculada con la ecuación (3-25).

En la tabla 4-5 se muestra la entropía para cada taxón en la cual se aprecia que en las redes metabólicas de organismos con funciones biológicas mas complejas la entropía es mayor.

Redes Bipartitas Enzimas - Compuesto

A diferencia de las redes metabólicas centradas en enzimas en este tipo de redes el grado de entrada k_{in} de una enzima cuantifica el número de sustratos que cataliza y el grado de salida k_{out} el número de productos que produce. Se analiza con gran interés el comportamiento de la tupla $[1, 1]$ que expresa la situación de una enzima que cataliza un único sustrato y genera a partir de ella un único producto

En la figura 4-5 se muestra el histograma comparativo entre las redes bipartita de los organismos estudiados, en el cual se expresa el porcentaje de enzimas con tupla $[1, 1]$ dentro de la red, En este sistema la tupla $[1, 1]$ es también la de mayor presencia en los 13 organismos estudiados. En esta comparativa podemos observar que a diferencia de lo que ocurre en las redes centradas en enzimas, el porcentaje de prevalencia de la tupla $[1, 1]$ en todos los taxones es aproximadamente la misma, cuyo valor oscila entre un 50 % y un 60 %.

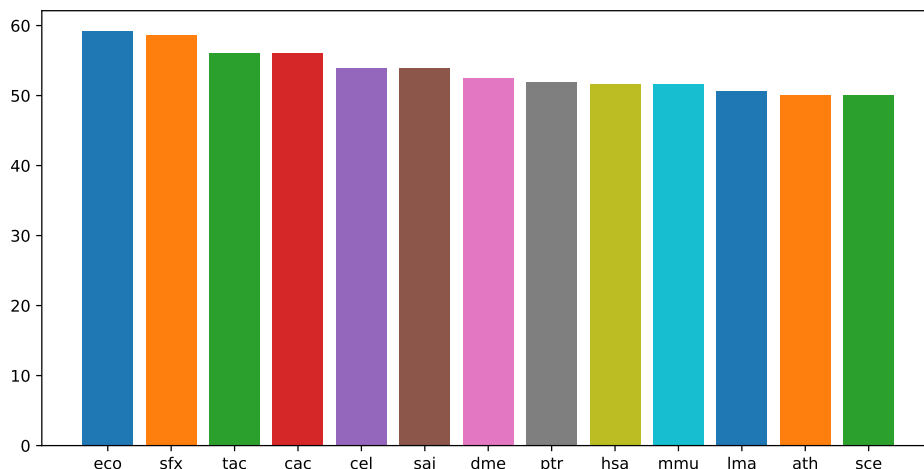


Figura 4-5: Histograma que indica el porcentaje de nodos con tupla $[k_{out}, k_{in}] = [1, 1]$ presentes en las redes metabólicas bipartitas de los organismos estudiados.

Al igual en el caso de las redes metabólicas centradas en enzimas, para cada una de las redes metabólicas bipartitas, existe una distribución propia de las tuplas $[k_{out}, k_{in}]$, la cual en el

mismo sentido refleja diferencias en los procesos que se llevan a cabo en el metabolismo. Para comparar esta distribución de tuplas entre los diferentes taxones, aplicamos la entropía como una medida de la dispersión o variabilidad de las tuplas dentro de las redes metabólicas bipartitas.

Tabla 4-6: Tabla de entropías asociadas con la dispersión de las tuplas $[k_{out}, k_{in}]$ en las redes metabólicas bipartitas. En gris los taxones del dominio eucariota.

Org	Entropía
hsa	1,94
ptr	1,92
mmu	1,91
lma	1,86
ath	1,84
cel	1,81
dme	1,80
sce	1,80
tac	1,71
sai	1,71
sfx	1,61
cac	1,59
eco	1,58

4.2.2. Discusión

En el estudio de la prevalencia de la tupla $[K_{out}, K_{in}] = [1, 1]$ en las redes metabólicas centradas en enzimas, observamos que ésta difiere en su proporción en diferentes organismos, estos resultados sugieren que en organismos eucariotas entre los cuales se encuentran organismos con funciones biológicas altamente sofisticadas y complejas como el Humano y el Chimpancé, se tiene una prevalencia mucho menor de la tupla $[K_{out}, K_{in}] = [1, 1]$ en comparación con organismos más simples como la Arquea *Sulfolobus solfataricus*. También observamos una mayor heterogeneidad de las tuplas de acuerdo a la entropía asociada con la distribución, en organismos eucariotas sofisticados en comparación con organismos procariotas con redes metabólicas de tamaño similar. De estos resultados se puede decir que existen diferencias significativas entre los organismos, expresadas en el marco de la variable estructural aquí definida, que corresponde con una correlación positiva del 91 % con las distancias evolutivas relativas entre los organismos eucariotas y el taxon *Homo sapiens* (hsa), y podría aportar información filogenética relevante. Para el cálculo de la correlación se utiliza el coeficiente

de Pearson entre los valores obtenidos de entropía y las distancias evolutivas relativas entre los organismos del dominio eucarita y el taxon hsa, medida como la diferencia entre ramas en el árbol filogenético.

Lo obtenido en las redes centradas en enzimas en cuanto a la prevalencia de la tupla $[K_{out}, K_{in}] = [1, 1]$ difieren mucho con los resultados de las redes metabólicas bipartitas, en éstas se observa que para los 13 organismos estudiados la prevalencia de tupla $[1, 1]$ tiene un porcentaje aproximadamente igual que oscila entre el 50 % y el 60 %, esta medida no expresa diferencias significativas que permitan correlacionarla con las distancias evolutivas de los organismos, mas sin embargo dado que en algún sentido la prevalencia de esta tupla está relacionada con el porcentaje de enzimas con una alta e especificidad, es posible afirmar que en las redes metabólicas las enzimas de este tipo son las mas relevantes.

En el análisis de la entropía asociada a la distribución de las tuplas en las redes bipartitas, sí existen diferencias significativas entre los organismos, con una correlación del 65 % con las distancias evolutivas relativas entre los organismos eucariotas y el taxon Homo sapiens. En la tabla [4-5] ordenada de mayor a menor podemos observar que los taxones estudiados del dominio Eucariota se encuentran en la parte superior de la tabla, por lo que esta medida podría aportar información filogenética relevante respecto a las diferencias estructurales del metabolismo en los diferentes dominios.

En la figura [4-6] se muestran dos redes metabólicas centradas en enzimas que ilustran dos situaciones hipotéticas que nos permiten discutir e interpretar en un sentido los resultados obtenidos de la prevalencia de la tupla $[K_{out}, K_{in}] = [1, 1]$, expresada de forma diferente en las redes. Suponemos que un camino primitivo es un metabolismo simple en el cual las enzimas se encuentran relacionadas por medio de un sólo metabolito con una única enzima, en un proceso análogo al de las vías metabólicas, en el cual todas las enzimas trabajan en una secuencia lineal para obtener un único producto principal, en este tipo de red la prevalencia de la tupla $[1, 1]$ es alta, en este caso es aproximadamente del 60 %. Por otra parte hablamos de una camino sofisticado como una red metabólica cuyas enzimas presentan un número mayor de interacciones con otras enzimas, en la cual no se habla únicamente de un proceso de transformacional lineal sino de la obtención por medio de rutas alternas de productos que pueden ser importantes, en este caso la prevalencia de la tupla $[1, 1]$ es inferior a 10 %. Es evidente que ambos procesos metabólicos son diferentes y que su estructura topológica tiene implicaciones por ejemplo en su robustez y resiliencia. En este sentido y a la luz de estas consideraciones, se construye la hipótesis de que una prevalencia menor de la tupla $[K_{out}, K_{in}] = [1, 1]$ está relacionada con una mayor sofisticación y complejidad en los procesos metabólicos, situación que expresa correlación con los procesos y las distancias evolutivas.

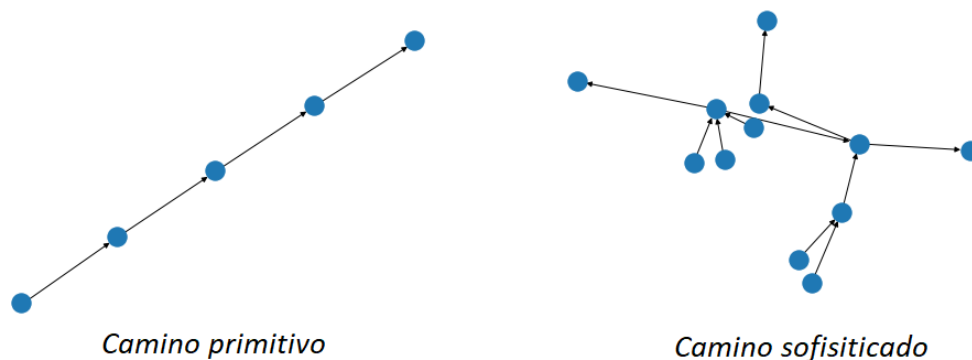


Figura 4-6: Redes metabólicas centradas en enzimas con diferente porcentaje de prevalencia de la tupla $[k_{out}, k_{in}] = [1, 1]$.

4.2.3. La diferencia de grado $k_{out} - k_{in}$

En general en el metabolismo el grado de entrada k_{in} y el grado de salida k_{out} se encuentran relacionados, Esto se debe a que la enzima necesita obtener un sustrato para posteriormente catalizarlo como parte del proceso de transformación de los alimentos dentro de la célula. Una forma de relacionar ambas medidas es mediante la diferencia de grado $k_{out} - k_{in}$, esta nueva medida basada en el grado permite clasificar las enzimas en dos grupos:

- Equilibradas: si $k_{out} - k_{in} = 0$. Es de esperar que la mayor parte de las enzimas tengan procesos equilibrados en el sentido que interactúan con un numero igual de enzimas de conectividad de entrada como de salida.
- Desequilibradas: si $k_{out} - k_{in} \neq 0$. En este categoría las enzimas pueden tener una diferencia positiva o negativa. Debido a que la red metabólica estudiada es centrada en las enzimas, es interesante el desequilibrio, dado que reúne a la mayoría de enzimas que catalizan alimentos sin transformaciones previas dentro del metabolismo y las que expulsan compuestos del sistema.

En Redes metabólicas centradas en enzimas

Una forma de visualizar y comparar el comportamiento de la diferencia de grado para un grupo de redes metabólicas es mediante sus distribuciones de probabilidad. En la Figura [4-7] se compara la distribución de probabilidad de la diferencia de grado para un grupo de taxones representativos de los dominios Eucariota, Procariota y Arquea. En ésta se puede observar que las distribuciones tienen un pico muy agudo en cero, lo que indica una mayor probabilidad de encontrar enzimas equilibradas en todas las redes metabólicas comparadas,

también que el desequilibrio es altamente simétrico en los organismos como se puede contrastar con los coeficientes de asimetría calculada que se muestran en la tabla [4-7] cuyo valores se encuentran entre $-0,5$ y $0,5$. También se observa una diferencia notable en la forma de las distribución entre organismos de diferentes dominios que se pueden relacionar a las entropías calculadas que se indican en la tabla [4-7].

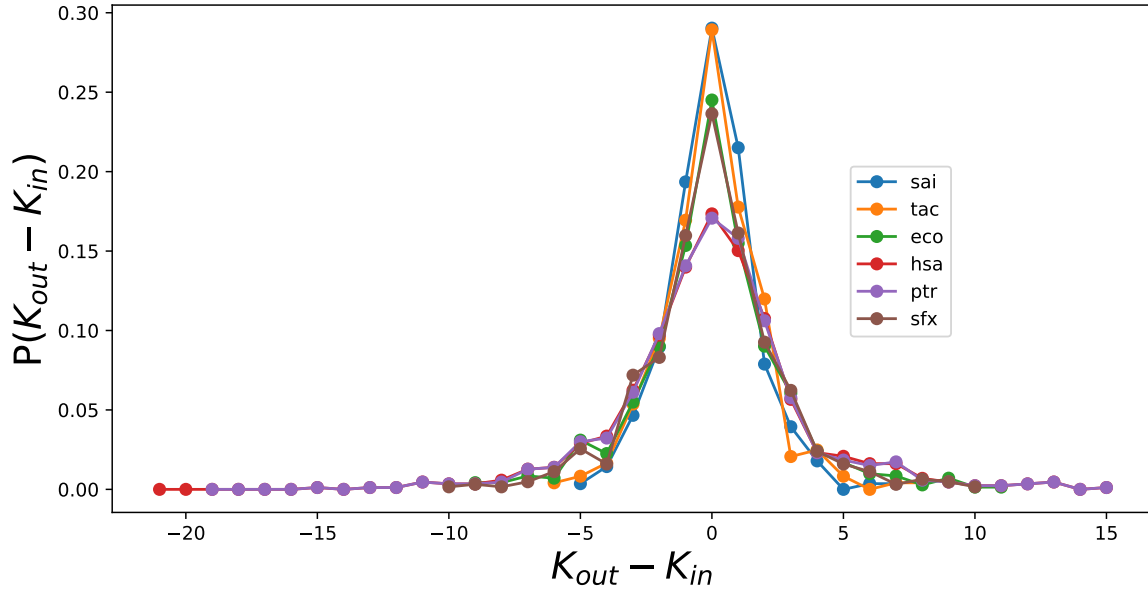


Figura 4-7: Distribución de la diferencia de grado de las redes metabólicas centradas en enzimas de los taxones estudiados mas representativos de cada dominio.

En la Figura [4-8] se muestra la distribución de la diferencia de grado para organismos con redes metabólicas con un numero de nodos aproximadamente iguales y de dominio diferentes, en esta se puede observar que organismos eucariotas como *Strongylocentrotus purpuratus* con 690 nodos conocidos tiene una distribución menos aguda en $k_{out} - k_{in} = 0$ comparado con organismos procariota como la *Escherichia coli* K-12 MG1655 cuya red metabólica tiene 682 nodos conocidos.

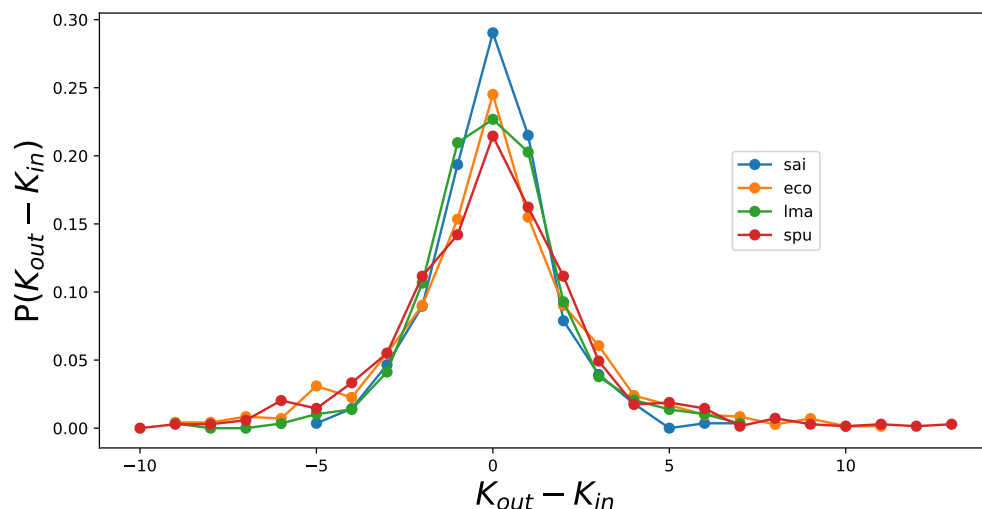


Figura 4-8: Distribución de la diferencia de grado de redes metabólicas centradas en enzimas de taxones con un número similar de enzimas y de dominios diferentes.

Tabla 4-7: Coeficiente de Asimetría de la distribución de diferencia de grado $k_{out} - k_{in}$, para las redes metabólicas centradas en enzimas, y la entropía asociada con la el número de estados accesibles y su dispersión en las distribución. En gris los taxones del dominio eucariota.

Org	Asimetria	Org	Entropia
ath	0,365	hsa	3,70
dme	0,286	ptr	3,68
sai	0,153	mmu	3,63
tac	0,146	eco	3,35
cac	0,130	sfx	3,31
mmu	0,128	ath	3,29
eco	0,117	cel	3,28
cel	0,094	dme	3,26
sfx	0,004	sce	3,22
ptr	-0,015	cac	3,04
sce	-0,034	lma	2,99
hsa	-0,081	tac	2,90
lma	-0,288	sai	2,75

En Redes metabólicas bipartitas

De forma análoga a la sección anterior, se analiza la diferencia de grado por medio de su distribución para las redes metabólicas bipartitas, tomando únicamente las enzimas de la red. En la figura [4-9] se muestran las distribuciones de probabilidad de la diferencia de grado para un grupo de taxones representativos de cada dominio, en esta se puede observar que, al igual que en las redes metabólicas centradas en enzimas, las distribuciones tienen un pico muy agudo en cero relacionada en el mismo sentido con una mayor probabilidad de encontrar enzimas equilibradas. Aunque a diferencia de los resultados anteriores en las redes bipartitas algunas distribuciones presentan un grado alto de asimetría (tabla [4-8]), pero unas distribuciones sin diferencias significativas con una relativa baja entropía y un valor aproximadamente igual (tabla [4-8]).

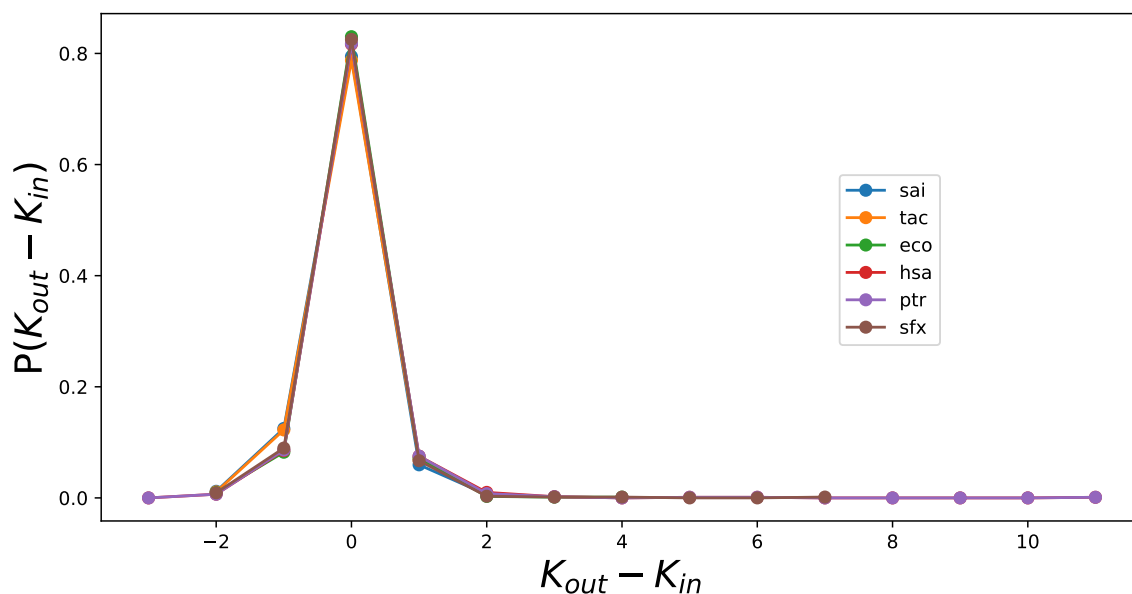


Figura 4-9: Distribución de probabilidad de la diferencia de grado de las redes metabólicas bipartitas de los taxones estudiados mas representativos de cada dominio.

Tabla 4-8: Coeficiente de Asimetría de la distribución de diferencia de grado $k_{out} - k_{in}$, para las redes metabólicas bipartitas, y la entropía asociada con la el número de estados accesibles y su dispersión en las distribución. En gris los taxones del dominio eucariota.

Org	Asimetria	Org	Entropia
ptr	5,86	tac	0,733
hsa	5,75	sce	0,718
sfx	2,97	sai	0,711
mmu	2,09	hsa	0,692
ath	0,35	lma	0,692
eco	0,29	ath	0,691
lma	-0,24	ptr	0,685
cel	-0,24	dme	0,677
sce	-0,50	mmu	0,677
tac	-0,75	cel	0,660
sai	-0,84	sfx	0,656
dme	-1,00	eco	0,635
cac	-1,02	cac	0,624

4.2.4. Discusión

En la distribución de la diferencia de grado $K_{out} - K_{in}$ en las redes metabólicas centradas en enzima, la entropía tiene una interpretación que corresponde tanto con el numero de estados accesibles en el sistema como a la dispersión que presenta la distribución. En la tabla [4-7] se ordenan los organismos de tal forma que la entropía sea máxima en la parte superior de la tabla, en esta se puede ver que organismos como el Humano y el Chimpancé se encuentran al principio de la tabla, precedido por organismos de sofisticación y tamaño metabólico creciente, esta medida de entropía presenta una correlación negativa de 91 % con las distancias evolutivas relativas al taxon Homo sapiens (hsa) y presenta diferencias significativas, por lo se podría obtener información filogenética apartar de ella. A la luz de estos resultados se formula como hipótesis que la evolución de las redes metabólicas va a favor de aumentar esta entropía, transformando la distribución es una más gaussiana. En cuanto a la forma de la distribución el coeficiente de asimetría en todos los organismos indica que estas son distribuciones aproximadamente simétricas, con una baja correlación negativa inferior al 13 % con las distancias evolutivas.

En el caso de las redes bipartitas la información asociada a la asimetría de las distribuciones, toma mas relevancia, cuyas diferencias son significativamente mas grandes y se tiene un coeficiente de correlación positiva de un 80 % con relación a las distancias evolutivas relativas,

este resultado sugiere que esta medida podría contener información filogenética importante. Como se puede observar, las distribuciones con un coeficiente de asimetría más grande tiene un sesgo positivo, el cual está relacionado en el metabolismo con enzimas cuyo grado k_{in} es siempre menor que el grado k_{out} , lo que indica la situación en la cual se catalizan un número menor de sustratos que productos posibles generados. En este sentido, parece que las presiones evolutivas propician la adición de enzimas en el metabolismo cuya diferencia de grado $k_{out} - k_{in}$ sea positiva. Por otra parte, la entropía asociada a esta distribución no presenta diferencias significativas que se pueden asociar a correlaciones evolutivas claras, con un coeficiente de correlación inferior al 24 % respecto a las distancias evolutivas relativas al taxón *Homo sapiens*.

4.3. Comparación de las dinámicas de las Redes metabólicas

En esta sección de resultados, se hace un estudio comparativo de la distribución de grado de las redes metabólicas centradas en enzimas de una muestra de 10 taxones representativa de cada dominio.

4.3.1. Distribución de grado nodal

Para comparar el grado nodal tanto de entrada como de salida de las redes metabólicas centradas en enzimas de los organismos estudiados se utiliza la distribución de grado. En las figuras [4-10] y [4-11] se muestran las acumulativas complementarias de las distribuciones de probabilidad de grado en escala semi-logarítmica, de organismos representativos de cada dominio y de una muestra de taxones del dominio eucarita respectivamente. Podemos observar que en estos organismos su distribución de probabilidad de grado exhiben un comportamiento exponencial, que nos sugiere que las redes metabólicas tienen este tipo de topologías, donde:

$$P(k) = Ae^{-\alpha k} \quad (4-1)$$

En la tabla [4-9] se muestran los parámetros del ajuste exponencial, su error cuadrático medio y el estadístico R^2 para cada uno de los organismos comparados. Adicionalmente en estas tablas se calcula la entropía asociada a estas distribuciones y se muestran una serie de valores de referencias que serán útiles en la discusión de estos resultados.

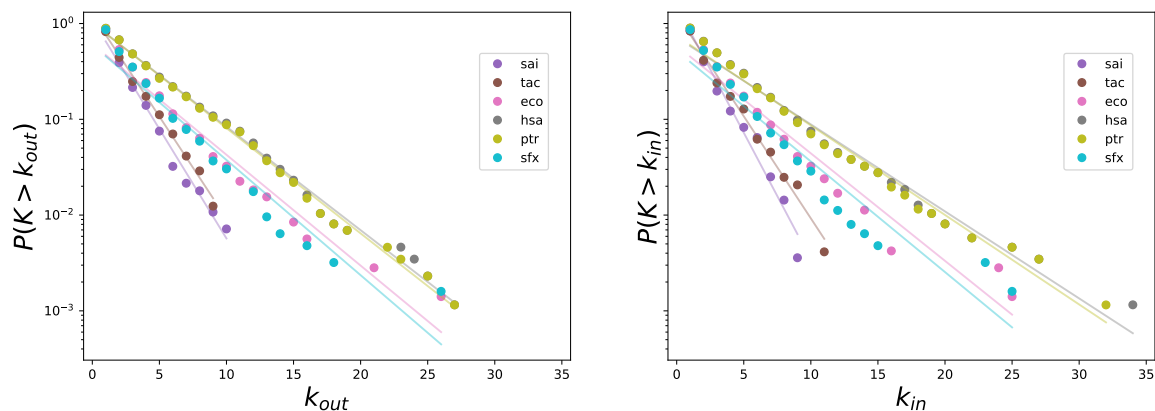


Figura 4-10: Distribución de grado k_{out} imagen izquierda y k_{in} imagen derecha, de las redes metabólicas de un grupo de taxones representativos de cada uno de los dominios Eucariota, Procariota y Arquea.

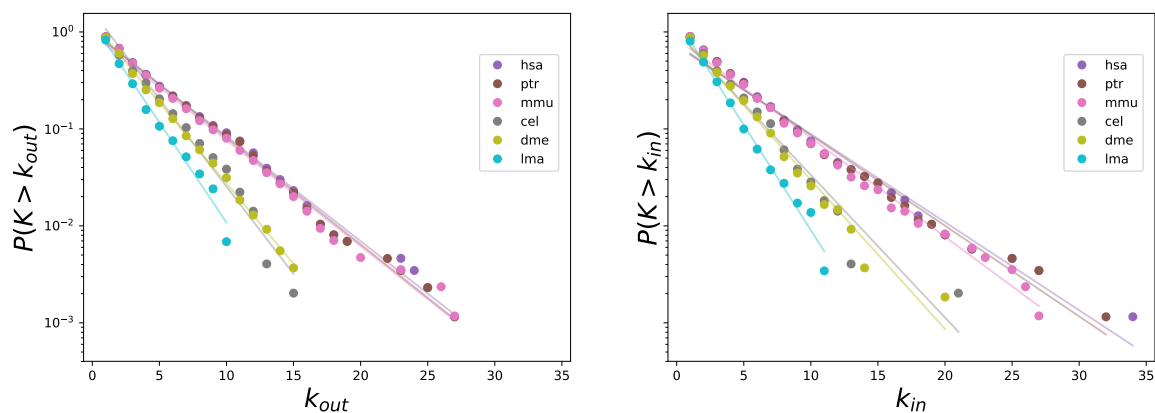


Figura 4-11: Distribución de grado k_{out} imagen izquierda y k_{in} imagen derecha, de las redes metabólicas de un grupo de taxones representativos de cada uno de los dominios Eucariota, Procariota y Arquea.

En la Tabla [4-10], se muestran las divergencias de Kullback-Leibler, calculadas con la ecuación (3-26), de las distribuciones de grado in y out de los taxones estudiado entre el organismos de referencia hsa. También se muestran las divergencias calculadas entre las distribuciones de probabilidad in y out para cada uno de los organismos.

Tabla 4-9: Parámetros de ajuste para las distribuciones exponenciales de grado k_{in} y k_{out} .
En gris los taxones del dominio eucariota.

Org	Distribución de grado k_{out}					Distribución de grado k_{in}					valores de referencia				
	A	α	ECM	R^2	S	A	α	ECM	R^2	S	\bar{k}	$1/\bar{k}$	$\bar{\alpha}$	α_{in}	α_{out}
hsa	2,81	0,249	0,027	0,992	3,74	2,21	0,209	0,067	0,978	3,78	3,82	0,262	0,229	0,209	0,249
ptr	2,83	0,253	0,018	0,995	3,69	2,26	0,215	0,055	0,982	3,74	3,77	0,265	0,234	0,215	0,253
mmu	2,69	0,252	0,036	0,990	3,56	2,47	0,235	0,033	0,990	3,63	3,71	0,270	0,243	0,235	0,252
eco	0,21	0,266	0,127	0,960	2,66	2,03	0,259	0,185	0,945	2,61	2,66	0,376	0,262	0,259	0,266
sfx	2,08	0,277	0,217	0,936	2,51	1,94	0,266	0,320	0,905	2,52	2,54	0,394	0,271	0,266	0,277
dme	3,47	0,383	0,004	0,999	2,56	2,93	0,354	0,084	0,974	2,54	2,75	0,364	0,368	0,354	0,383
cel	4,40	0,415	0,080	0,974	2,65	2,84	0,338	0,172	0,945	2,59	2,84	0,352	0,377	0,338	0,415
lma	3,49	0,475	0,041	0,978	1,99	3,71	0,500	0,041	0,984	1,97	2,06	0,485	0,487	0,500	0,475
tac	3,50	0,491	0,011	0,993	1,89	3,52	0,491	0,030	0,987	1,93	1,96	0,510	0,491	0,491	0,491
sai	3,25	0,526	0,034	0,985	1,67	4,21	0,609	0,080	0,969	1,65	1,76	0,568	0,568	0,609	0,526

Tabla 4-10: Divergencias de Kullback-Leibler (D_{KL}) respecto a las distribución de probabilidad de grado in y out del taxon Homo sapiens (hsa). Divergencias de Kullback-Leibler entre las distribuciones de probabilidad in y out de los organismos.

Divergencias de Kullback-Leibler respecto a las distribución hsa				Entre distribuciones out e in	
Distribución de grado k_{out}		Distribución de grado k_{in}			
Org	D_{KL}	Org	D_{KL}	Org	D_{KL}
hsa	0,000	hsa	0,000	hsa	0,0333
ptr	-0,047	ptr	-0,046	ptr	0,0319
mmu	-0,122	mmu	-0,114	cel	0,0178
cel	-0,610	cel	-0,581	mmu	0,0172
dme	-0,680	dme	-0,663	lma	0,0145
eco	-0,770	eco	-0,746	tac	0,0136
cac	-0,851	cac	-0,819	dme	0,0129
lma	-0,861	sai	-0,847	sai	0,0105
tac	-0,862	lma	-0,856	eco	-0,0022
sai	-0,862	tac	-0,886	cac	-0,0134

4.3.2. Discusión

Las distribuciones de grado k_{out} y k_{in} en las redes centradas en enzimas para los 10 organismos siguen una ley exponencial con $\alpha \sim 0,209 - 0,500$ en los organismos eucariotas y $\alpha \sim 0,491 - 0,609$ en las arqueas, ambos grupos de organismos con fuertes correlaciones $R^2 \sim 0,978 - 0,990$, mas sin embargo las bacterias cuyo $\alpha \sim 0,259 - 0,277$ presentan una correlación más baja con $R^2 \sim 0,905 - 0,960$, mas sin embargo todas las distribuciones son altamente exponenciales, cuyos exponentes tienen una correlación (correlacion de Pearson) positiva del 96 % con lo obtenido en ACP para los taxones eucariotas. También α se relacionan de forma muy cercana con el inverso del grado promedio $1/\bar{k}$, por lo que aproximadamente se cumple la relación (4-2), aunque la diferencia entre ambas medidas podrían aportar información relevante del sistema como lo mencionaremos mas adelante.

$$P(k) \approx e^{-\frac{k}{\bar{k}}} \quad (4-2)$$

En (4-2) k hace referencia al grado k_{out} o k_{in} . Según la teoría de las redes complejas, esta topología exponencial indica que estas redes no son estáticas, sino evolutivas pero cercanas al equilibrio, cuyo proceso de evolución involucra la unión de vértices al azar, (sin ninguna preferencia, pero con cantidades conservadas.)[20]. Otras redes en la naturaleza con topología exponencial, son la red de transporte urbano dentro de una ciudad y la red mundial de transporte marítima (ejemplos que pueden ser una buena analogía al sistema de rutas metabólica), estudiados por Weibing Deng, Wei Li y Qiuping en el 2011 [25], quienes proponen que la evolución de estos sistemas de transporte se basa en dos mecanismos: el crecimiento y el adjunto aleatorio adyacente, es decir que la adición de un nuevo vértice A a otro B en la red implica una adición aleatoria también de A a los nodos adyacentes al vértice B . Este estudio nos permite dar una interpretación mas amplia de los resultados, y construir la hipótesis de que las redes metabólicas crecen con la adición de nuevas enzimas a otra sin preferencias y luego a enzimas cercanas a está. Esta hipótesis queda por estudiar en detalle en un trabajo futuro.

La entropía asociada tanto a las distribuciones de grado k_{in} como k_{out} de los organismos eucariotas tiene una correlación de Pearson negativa del 94 % con las distancias evolutivas de estos taxones, lo cual indica que este aspecto dinamico de las redes metabólicas puede contener información filogenética relevante, y en la que se podría interpretar la evolución como un proceso de maximización de esta entropía.

Cuando las distribuciones de probabilidad de una red en estado estacionario es exponencial, el sistema puede modelarse de manera análoga a un gas de partículas en equilibrio térmico [26]. Debido a que las distribuciones de grado k_{in} como k_{out} son altamente exponenciales para todos los organismos, podemos modelarlas aplicando una propiedad de maximización de la entropía. Sabemos que el sistema metabólico evoluciona en una escala de tiempo grande,

por lo que suponemos que está en quasi-equilibrio. De esta manera, podemos construir un modelo estadístico a partir de un ensamble canónico de grafos [20](cuyos valores esperados de los observables coincidan con los de las redes metabólicas reales), el cual se desarrollara como un trabajo a futuro. Nuestra hipótesis es que las diferencias entre los valores α y el observable $1/\bar{k}$ en las redes metabólicas, está relacionada con la existencia de ligaduras en el sistema, siendo una de ellas $\langle \bar{k}_{out} - \bar{k}_{in} \rangle = 0$, lo que implicaría:

$$P(k) \sim e^{-\lambda \langle \bar{k}_{out} - \bar{k}_{in} \rangle} \quad (4-3)$$

Siendo λ el multiplicado de Lagrange asociada con el valore esperado de la diferencia de grado promedio $\langle k_{out} - k_{in} \rangle$. Eventualmente el sistema puede tener un conjunto de ligaduras diferente para organismos distantes evolutivamente.

Por otra parte, los resultados obtenidos del calculo de la divergencia de Kullback-Leibler (D_{KL}), muestran la estrecha similitud entre las distribuciones de probabilidad de los organismos evolutivamente cercanos, correlacionandose muy bien con las distancias evolutivas relativas al taxón hsa. También es un resultado interesante que los organismos del dominio eucariota, en general presentan distancias D_{KL} , entre las distribuciones de grado k_{in} y k_{out} , más grandes respecto a los demás organismos, esto nos indica que probablemente, en estos organismos, son acentuadas las diferencias entre las dinámicas de grado k_{in} y k_{out} , en las redes metabólicas.

5 Conclusiones e hipótesis

En este capítulo se reúnen todas las conclusiones e hipótesis desarrolladas a partir de la discusión de los resultados obtenidos.

5.1. Conclusiones

- Los organismos estudiados poseen redes metabólicas con topologías distintas, y sus diferencias topológicas son significativas y correlacionadas con las distancias evolutivas entre ellas.
 - En las redes centradas en enzimas, la prevalencia de la tupla $[k_{out}, k_{in}] = [1, 1]$, asociada a la relación enzimática más frecuente, está significativamente correlacionada con las distancias evolutivas entre organismos.
 - Los resultados obtenidos de las entropías asociadas a las distribuciones de probabilidad de las tuplas $[k_{out}, k_{in}]$ y las diferencias de grado $k_{out} - k_{in}$ en las redes centradas en enzimas, indican dinámicas más diversas en el metabolismo de organismos más sofisticados.
-
- Los organismos estudiados tienen redes metabólicas con topologías distintas.
 - Las redes metabólicas de los 13 organismos estudiados presentan diferencias topológicas significativamente correlacionadas con las distancias evolutivas entre ellos.
 - Los resultados obtenidos en la heterogeneidad y dispersión asociadas con las medidas de entropía de las distribuciones de tuplas y diferencia de grado $k_{out} - k_{in}$. La diversidad que reflejan las dinámicas
 - En las redes centradas en enzimas, el estudio de las prevalencias de las tuplas $[k_{out}, k_{in}]$, asociada a las relaciones enzimáticas más frecuentes, indica que la presencia de la tupla $[1, 1]$ y la diversidad de tuplas tiene una correlación significativa con las distancias evolutivas.
 - La entropía asociada con la heterogeneidad de nodos con diferencias de grado $k_{out} - k_{in}$ dado, expresa una correlación significativa con las distancias evolutivas, y con distribuciones de probabilidad aproximadamente simétricas en las redes centradas en enzimas.

- El coeficiente de asimetría asociado a las distribuciones de probabilidad de la diferencias de grado $k_{out} - k_{in}$, muestra que algunos taxones sofisticados tienen un sesgo positivo grande, y que difiere significativamente entre los organismos, lo que indica que esta medida probablemente contiene información filogenética relevante.
- Las redes metabólicas centradas en enzimas de los organismos estudiados presentan topologías fuertemente exponenciales, indicativa de conectividad evolucionada al azar.
- Los organismos eucariotas mas sofisticados presentan distancias D_{KL} significativamente mas grandes entre las distribuciones de grado k_{out} y k_{in} , que organismos mas simples, indicando una mayor divergencia en las dinámicas asociadas al grado k_{out} y k_{in} en organismos biológicamente mas sofisticados.
- Las redes metabólicas pueden ser modeladas a partir una imagen mecánico estadística, como sistemas complejos evolutivos en estado quasi-estático, con una dinámica emergente asociada a un principio de maximización de la entropía asociada a la distribución de probabilidad de grado in y grado out, sujeto a ligaduras; con un ensamble estadístico de grafos aleatorios.

5.2. Hipótesis

- En las redes metabólicas centradas en enzimas una menor prevalencia de la tupla $[k_{out}, k_{in}] = [1, 1]$ indica procesos metabólicos mas sofisticados y complejos.
- La presión evolutiva propicia la adición de enzimas que catalizan un mayor número de productos a partir de un número menor de sustratos.
- La evolución de las redes metabólicas favorece la heterogeneidad de nodos con interacciones estructurales más diversas.
- Las redes metabólicas evolucionan a partir de la adicción de nuevas enzimas, con las siguientes características; que al unirse al metabolismo lo hace sin preferencia alguna por una o un grupo de enzimas, y que los nuevos vínculos que esta establece en la red son con enzimas adyacentes a sus vecinas.

6 Anexos

En esta sección se muestran resultados adicionales, que aporta información sobre las diferencias de los sistemas metabólicos de los organismos estudiados.

6.1. Prevalencia de las diferentes clases de enzimas en las Redes metabólicas

El sistema metabólico consta de enzimas que llevan acabo tipos de reacciones bioquímicas diferentes, y de acuerdo a las características de estas, son clasificadas en los siguientes grupos; Oxidoreductansas, Transferasas, Hidrolasas, Liasas , Isomerara y Ligasas. Reacciones llevadas acabo por las enzimas tipo ligasas, por ejemplo, hacen que un grupo de sustratos reaccionen formando un solo compuesto, es por tal razón que probablemente presentan dinámicas particulares dentro de la Red.

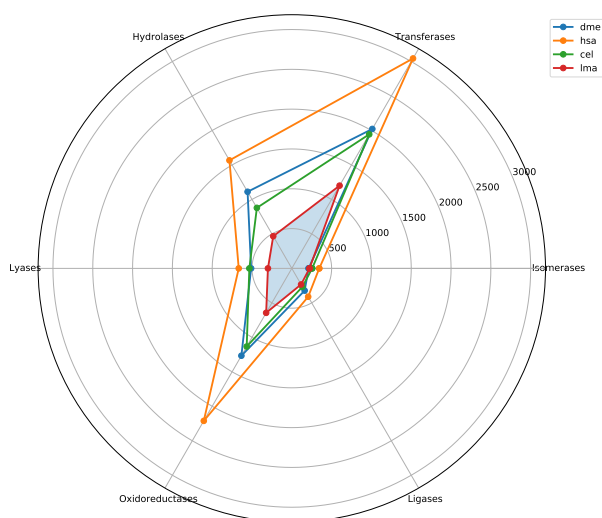


Figura 6-1:

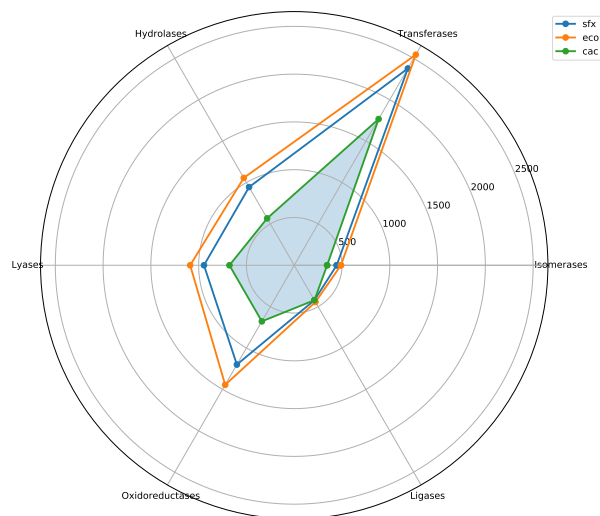


Figura 6-2:

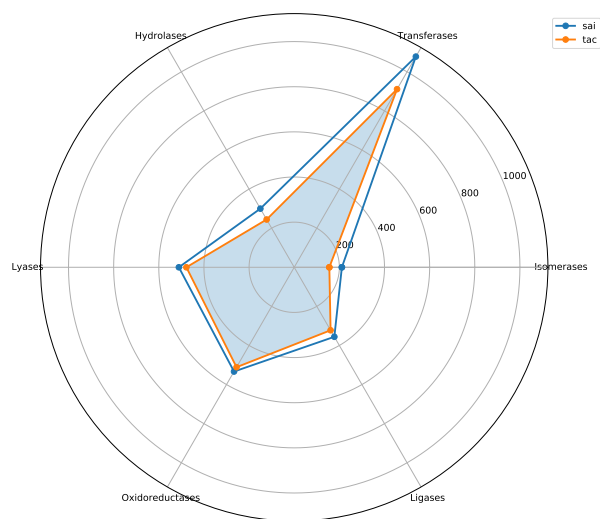


Figura 6-3:

En la figura **6-1** se muestran la prevalencia de los diferentes tipos de enzimas dentro de la red metabólica de los organismos eucariotas. En las figuras **[6-2]** y **[6-3]** se muestran las

prevalencias de los diferentes tipo de enzimas para los organismos del dominio bacteria y arquea respectivamente. En Todas estas figuras podemos apreciar que las distribuciones de enzimas de cada clase en el sistema metabólico son muy cercanas en organismos del mismos dominio, y diferentes entre organismos de diferentes dominios como se muestra en la figura [6-4].

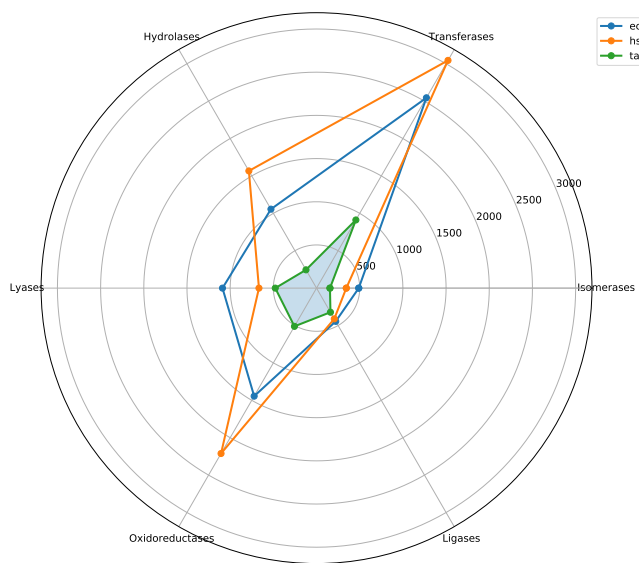


Figura 6-4:

Cualitativamente los cambios que se aprecian en la prevalencia de los tipos de enzimas entre organismos de un mismos dominio son suaves y tienen aproximadamente la misma estructura para organismos con tamaños diferentes, también notamos que los diagramas de metabolismos mas pequeños se encuentran completamente contenidos dentro del diagrama de redes metabólicas mas grandes, a diferencia de la comparativa entre organismos de diferentes dominios como se aprecia en la figura [6-4], en los cuales las diferencias son tan significativas que las área de sistemas metabólicos mas pequeños no están completamente contenidos en los de organismos mas grandes.

Bibliografía

- [1] Libia Herrero Uribe. ¿ qué es la vida?¿ la ciencia, se atreve a definirla? *Diálogos Revista Electrónica*, 7(1):1–35, 2006.
- [2] Enrique Meléndez Hevia. *La evolución del metabolismo: hacia la simplicidad*. Eudema, 1992.
- [3] Donald Voet and Judith G Voet. *Bioquímica*. Ed. Médica Panamericana, 2006.
- [4] Hong-Wu Ma and An-Ping Zeng. Phylogenetic comparison of metabolic capacities of organisms at genome level. *Molecular phylogenetics and evolution*, 31(1):204–213, 2004.
- [5] Charles Darwin. El origen de las especies charles darwin. 2009.
- [6] Juan J Morrone. *Homología biogeográfica: las coordenadas espaciales de la vida*, volume 37. UNAM, 2004.
- [7] Pablo Vargas and Rafael Zardoya. *El árbol de la vida: sistemática y evolución de los seres vivos*. Vargas, Pablo, 2012.
- [8] Jorge Granados Zúñiga. La evolución de los sistemas bioquímicos complejos. *Revista de Filosofía de la Universidad de Costa Rica*, 57(147):57–69, 2018.
- [9] Santiago Ramírez. *Perspectivas en las teorías de sistemas*. Siglo XXI, 1999.
- [10] Maximino Aldana. Redes complejas. *Recuperado a partir de <http://www.fis.unam.mx/~max/English/notasredes.pdf>*, 2006.
- [11] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.
- [12] JM Berg. *Metabolismo: conceptos básicos y visión de conjunto*, chapter 15.1, pages 409–412. Editorial Reverté S.A., 2008.
- [13] Mónica Barboza. <https://co.pinterest.com/pin/538109855470933643/>, s.f.
- [14] Roger Guimerà and Marta Sales-Pardo. La promesa de las redes metabólicas, 2015.

- [15] José Leonardo Bocanegra Silva. *Análisis comparativo de las redes metabólicas de metatranscriptomas edáficos de la rizósfera de Solanum phureja en tres diferentes estados fenológicos.*, chapter 1.5. Trabajo de grado, Universidad Nacional de Colombia Facultad de Ciencias, 2015.
- [16] La evolución de la conectividad en redes metabólicas. *PLoS Biol.*
- [17] J Peretó. Origin and evolution of metabolisms. pages 270–287, 2011.
- [18] Xabier Barandiaran and Miguel Aguilera. Neurociencia y tecnopolítica: hacia un marco analógico para comprender la mente colectiva del 15m. *Tecnopolítica y M*, 15:163–211, 2015.
- [19] Albert-László Albert, Réka Albert, and Barabási. Mecánica estadística de redes complejas. *Reseñas de física moderna*.
- [20] Sergey N Dorogovtsev and Jose FF Mendes. Evolution of networks. *Advances in physics*, 51(4):1079–1187, 2002.
- [21] Diógenes Campos Romero. Elementos de mecánica estadística. 2006.
- [22] Jonathon Shlens. Notes on kullback-leibler divergence and likelihood theory. *Webseite (20. August 2007)* <http://www.snl.salk.edu/~shlens>, 2007.
- [23] Comparación de redes complejas: en defensa de lo simple. *New Journal of Physics*.
- [24] Carlos Lozares Colina and Pedro López-Roldán. El análisis de componentes principales: aplicación al análisis de datos secundarios. *Papers: revista de sociologia*, (37):031–63, 1991.
- [25] La distribución de grados exponencial en redes complejas: teoría de redes de no equilibrio, simulación numérica y datos empíricos. *Physica A: Mecánica estadística y sus aplicaciones*.
- [26] Mecánica estadística del dinero. *The European Physical Journal B-Condensed Matter and Complex Systems*.