

Redes metabólicas

Luceth Catherine Argote Radillo

Samuel Fernando De Dios Pérez

Lukas Morera Torres

Link del código fuente: [Google Colab](#)

1. Introducción.

Dentro del estudio biológico de los diversos sistemas de los seres vivos, el metabolismo es una de las áreas más importantes en este campo específico que relacionan estructuras biológicas. De este estudio se ha derivado la necesidad de representar los procesos metabólicos de una manera más sistémica y “entendible”, esto para poder ver con mayor claridad cómo se desarrollan ciertos procesos específicos.

Así es como se introducen las redes metabólicas, las cuales son una representación de los procesos químicos que se llevan a cabo dentro de los seres vivos, junto con los metabolitos, que son reactivos o compuestos químicos que forman parte de estas reacciones.

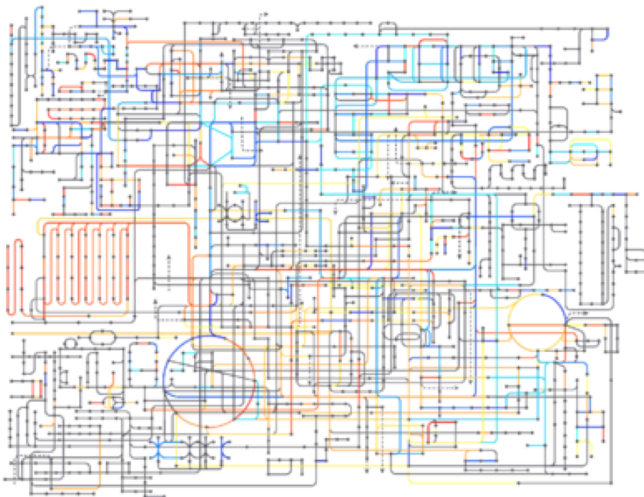


Figura 1: Representación gráfica de una red metabólica.

Las redes metabólicas han sido de gran utilidad para el estudio de diferentes organismos y tener una

mejor comprensión de ciertas reacciones del mismo, esto ya qué, si se analiza desde una perspectiva enteramente biológica, aumenta mucho la complejidad de comprensión de ciertas situaciones, mientras que desde una perspectiva sistémica y de redes se pueden comprender mejor diversos procesos que acontecen dentro de los seres vivos. [8]

2. Descripción del problema.

Dentro del estudio del metabolismo, se han encontrado múltiples reacciones que no se han podido explicar correctamente, esto debido a que hay vacíos en estas relaciones que ocasionan que estén incompletas, siendo los vacíos en las redes metabólicas uno de los principales dificultades al tratar estos sistemas. [2]

Es así como el problema se plantea hacia analizar si en una red metabólica específica, en este caso la de la bacteria *Escherichia coli*, se encuentran vacíos en los metabolismos en las diferentes reacciones que se puedan encontrar, para después proponer una solución distinta a la biológica y que aplique la teoría de grafos que permita llenar estos vacíos mediante metabolitos ya existentes en la red que generen una serie de reacciones que permitan llenar estos vacíos.

3. Objetivo general.

Realizar un análisis de centralidad de una red metabólica específica de la bacteria *Escherichia coli* por medio de la teoría de grafos usando diversos algoritmos, abstrayendo ciertas características de los mismos.

4. Objetivos específicos.

1. Modelar una red metabólica por medio de grafos.
2. Identificar y completar los vacíos (Gaps) que se puedan encontrar en la red metabólica.
3. Realizar un análisis de la red metabólica previo y posterior a la modificación de la misma.

5. Marco teórico.

- **Análisis de centralidad:** Es el estudio dentro de un grafo que revisa la importancia de cada vértice dentro del grafo, esto medido de acuerdo a distintos índices cuantitativos. Algunas de las medidas de centralidad más importantes son:
 - **Centralidad de grado:** Mide el grado de cada vértice dentro del grafo, es decir, cuantas aristas inciden en ese vértice.
 - **Centralidad de cercanía:** Calcula la suma o el promedio de las distancias mínimas de cada vértice con todos los demás dentro del grafo.
 - **Centralidad de intermediación:** Calcula las veces que aparece un mismo vértice en todos los caminos más cortos de otros vértices del grafo a los demás.
 - **Centralidad de valor propio:** Calcula la influencia de un nodo en un grafo mediante su principal valor propio.

Al trabajar sobre grafos dirigidos, es necesario definir algunas variantes para algunos de los métodos anteriores:

- **Centralidad de grado de entrada:** Mide el grado de entrada de cada vértice dentro del digrafo, es decir, cuantas aristas entran en ese vértice.
- **Centralidad de grado de salida:** Mide el grado de salida de cada vértice dentro del digrafo, es decir, cuantas aristas salen de ese vértice.
- **Centralidad de Katz:** Nombrada por su creador Leo Katz, es una variante de la centralidad por valor propio, con la

diferencia de que genera un valor generalizado para digrafos, teniendo en cuenta las salidas y entradas de cada nodo.

En el caso de la centralidad de cercanía e intermediación, funcionan igual en digrafos, solo teniendo en cuenta la dirección de las aristas al buscar los caminos pertinentes [10].

- **Red metabólica:** Es una representación de las interacciones bioquímicas que ocurren en un sistema biológico, en relación con las rutas y procesos metabólicos. En esta representación, los nodos representan metabolitos (moléculas) y las reacciones químicas, mientras que las uniones indican cómo los metabolitos participan en las diversas reacciones. Estas redes capturan cómo se transforman los nutrientes en productos, cómo se genera energía, cómo se sintetizan compuestos esenciales, lo que permite una mejor comprensión de cómo funcionan los sistemas biológicos a nivel molecular. [9]
- ***Escherichia coli* (*E. coli*):** Se va a trabajar sobre este ser vivo el cual es una bacteria gramnegativa (doble membrana) que se encuentra en el tracto gastrointestinal de algunos animales de sangre caliente. Además, *E. coli* es un organismo modelo en la investigación biológica debido a su facilidad de cultivo y estudio, y ha sido utilizado para investigar procesos biológicos y producir proteínas en biotecnología [7].

6. Modelamiento del problema.

Para modelar la red metabólica de manera eficiente, es necesario separar los vértices en dos tipos: Metabolitos y Reacciones, los cuales son los dos elementos principales a tener en cuenta en el modelamiento de una red metabólica, ya que ambos funcionan como las bases de la misma [3].

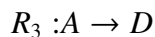
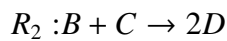
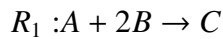
Por ende, es necesario representar dos tipos de escenarios, los cuales serían los metabolitos que participan en una reacción, y los metabolitos producidos por una reacción. Así, podemos ver que es necesario tener un orden en las uniones de cada elemento, ya que son escenarios distintos.

Teniendo en cuenta lo anterior, se decidió que la

manera óptima de representar la red metabólica es mediante un grafo bipartito dirigido, el cual cuenta con un conjunto de metabolitos y un conjunto de reacciones. De esta manera:

- Si unos metabolitos se conectan hacia una reacción, significa que ese grupo de metabolitos producen esa reacción.
- Si una reacción se conecta hacia unos metabolitos, significa que esa reacción produce esos metabolitos.

A manera de ejemplo de una red metabólica, planteamos el siguiente esquema de reacciones:



Estas reacciones se modelarían de la siguiente manera:

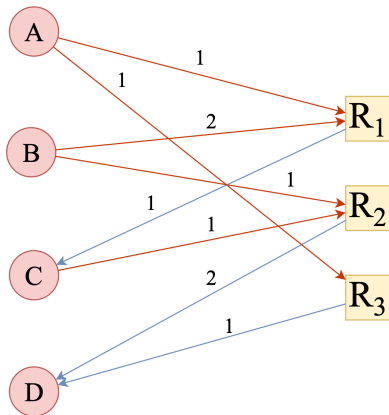


Figura 2: Modelamiento de una red metabólica.

En la Figura 2, Los metabolitos son A, B, C y D , mientras que las reacciones son R_1, R_2 y R_3 . Respecto a las aristas, el color rojo representa la unión de los metabolitos para crear una reacción, mientras que el azul representa el metabolito resultante de dicha reacción. Estos colores se implementan para una mejor comprensión del modelo, ya que el mismo se puede hacer bastante grande, por lo cual esta notación resulta muy útil. Por último, cada arista cuenta con un peso, el cual simboliza el número de moles necesarios para que la reacción se lleve a cabo.

7. Solución propuesta.

Para darle solución al problema descrito anteriormente se va a implementar la red metabólica en Python con la ayuda de diversas librerías que permitan resolver el problema a cabalidad.

Una vez que implementado el grafo correctamente, se procederá a obtener toda la información relevante del mismo con ayuda de las librerías correspondientes, y así mismo se obtendrá una visualización del grafo resultante para tener una idea de como se ve la red metabólica de la E.Coli.

Finalmente, se sacarán las conclusiones pertinentes con respecto a los datos y visualización de la red que se hayan obtenido, todo esto con base en la información planteada en el marco teórico, y así obtener un análisis completo de esta red metabólica. Cabe destacar que cualquier base de datos que se elija de *BiGG Models* puede modelarse en un grafo con las características que se requieran.

8. Implementación de la solución.

La solución del proyecto se realizó por medio del lenguaje de programación Python, esto mediante el uso de las siguientes librerías:

- **Networkx:** Sirve para implementar grafos dentro de Python, permitiendo definir diferentes características de los mismos como podrían ser las etiquetas, colores, direcciones, etc. Así mismo, también tiene la posibilidad de implementar ciertos algoritmos para obtener datos o subgrafos específicos de un grafo dado.
- **Matplotlib.pyplot:** Sirve para visualizar gráficamente el grafo, junto con todas las características que se hayan definido del mismo.
- **csv:** Sirve para que Python pueda leer bases de datos en formato CSV, formato en el que se encuentra la información de la red que queremos analizar.

Para empezar, se cargan las bases de datos relacionadas a la red metabólica de la bacteria E.Coli con ayuda de la librería *csv*. Se utilizan dos bases de datos principales:

- **Metabolitos:** Tiene en su interior todos las

enzimas químicas que producen las reacciones dentro de la red. Las mismas se identifican por un ID.

- **Reacciones:** Tiene en su interior todas las reacciones posibles entre los metabolitos de la red. Cada una tiene los metabolitos que necesita para ser producida y los metabolitos producidos por dicha reacción, así como la cantidad de moles de cada uno. También se representan con una ID.

Estas dos bases de datos se pasan a Python, y con base a la misma se arma un modelamiento por medio de la librería *Networkx* con las siguientes características:

- Es un grafo dirigido, donde los metabolitos que entran a las reacciones son los que las producen, mientras que los que salen de las reacciones son los productos de las mismas.
- Los metabolitos se representan como vértices de color rojo, y las reacciones como vértices de color azul.
- El grafo es bipartito, separándose en un grupo para metabolitos y otro para reacciones.

Una vez implementado con las características mencionadas, la visualización mediante la librería *Matplotlib.pyplot* es la siguiente:

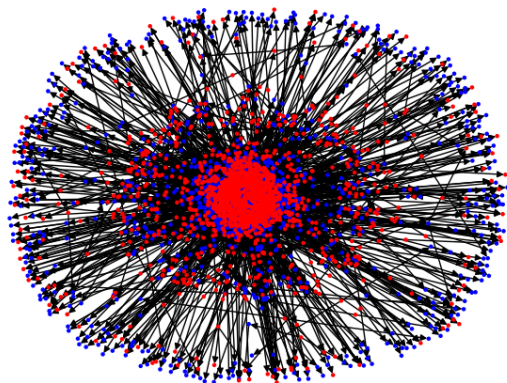


Figura 3: Modelamiento de la red completa.

El grafo visualmente no es muy entendible, ya que es un número bastante alto de metabolitos y reacciones, por lo cual no se puede ver muy claramente

todo lo que sucede visualmente. Sin embargo, para hacer el análisis de centralidad solo es necesario que el grafo esté implementado dentro de Python. Es importante recordar qué, para que el análisis de centralidad se pueda llevar a cabo, el grafo debe ser conexo, característica que, mediante Python, se verificó que no se cumplía en la red. Por ello, se decidió trabajar con dos subgrafos de la red: La componente fuertemente conexa máxima y la componente débilmente conexa máxima, teniendo así los grafos para el análisis de centralidad.

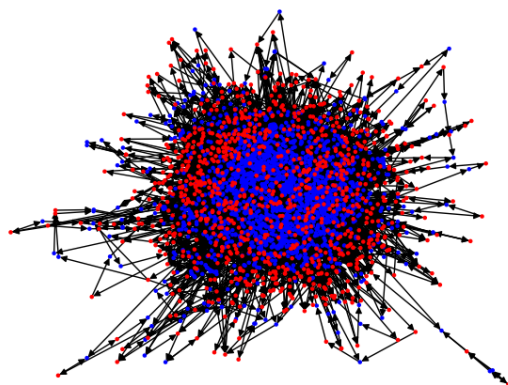


Figura 4: Modelamiento de la red fuertemente conexa.

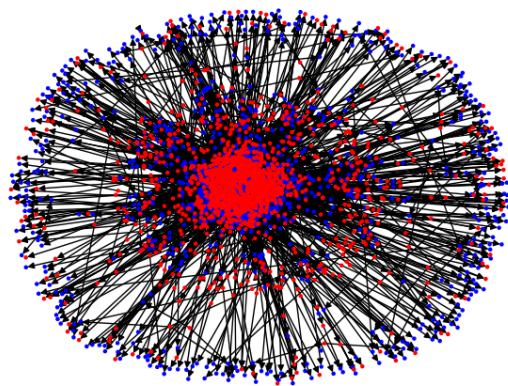


Figura 5: Modelamiento de la red débilmente conexa.

Como se puede apreciar, ambos grafos no son muy entendibles de manera visual. Por ende, por motivos estéticos y de comprensión, la visualización se hará con una porción más pequeña del grafo, esto para que se visualicen de mejor manera las características principales de la red.

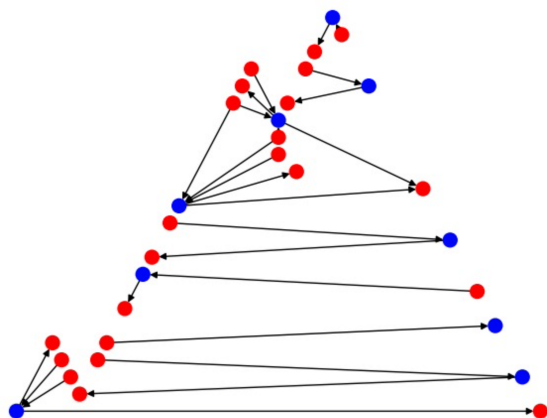


Figura 6: Modelamiento de una red simple.

9. Resultados obtenidos.

Por medio de la librería *Networkx* es posible obtener todas las características de centralidad de ambas redes, donde se obtuvieron los valores mas altos para cada método, separando el análisis para metabolitos y reacciones. Así, estos fueron los resultados obtenidos para cada una:

Red fuertemente conexas:

- **Centralidad de grado de entrada:** En metabolitos, los resultados fueron

1. *h_c*, con un grado de 0,2532
2. *adp_c*, con un grado de 0,0935
3. *pi_c*, con un grado de 0,0888

Por otro lado, para las reacciones fueron:

1. *ASNS 1*, con un grado de 0,0014
2. *CBPS*, con un grado de 0,0014
3. *CTPS 2*, con un grado de 0,0014

Respecto a la centralidad de grado de entrada, se puede interpretar como el índice de popularidad de un nodo dentro de la red [6]. Esto quiere decir qué, como es de entrada, los metabolitos mencionados son los que mas se producen por medio de las reacciones dentro de la red.

Respecto a las reacciones, las mencionadas son las que mas requieren metabolitos para que se lleven a cabo. Sin embargo, cabe destacar que todas tienen un grado de entrada similar, por lo que se puede concluir que no

varía mucho el número de metabolitos necesario de una a otra.

- **Centralidad de grado de salida:** En metabolitos, los resultados fueron

1. *h_c*, con un grado de 0,2532
2. *adp_c*, con un grado de 0,0935
3. *pi_c*, con un grado de 0,0888

Por otro lado, para las reacciones fueron:

1. *ASNS 1*, con un grado de 0,0014
2. *CBPS*, con un grado de 0,0014
3. *CTPS 2*, con un grado de 0,0014

Se puede apreciar que los resultados son análogos a los del grado de entrada. Y como el grado de salida tambien es un índice de popularidad de un nodo dentro de la red [6], se puede concluir qué, en el caso de los metabolitos, aquellos que son mas producidos, tambien son los que mas se emplean en las reacciones.

Por otro lado, las reacciones que mas emplean metabolitos tambien son las que mas producen metabolitos, viendose así una equivalencia entre producir y ser producido dentro de la red.

- **Centralidad de cercanía:** En metabolitos, los resultados fueron

1. *adp_c*, con un grado de 0,2668
2. *ppi_c*, con un grado de 0,2654
3. *h2o_c*, con un grado de 0,2553

Por otro lado, para las reacciones fueron:

1. *ADD*, con un grado de 0,2860
2. *GUAD*, con un grado de 0,2850
3. *CSND*, con un grado de 0,2849

Respecto a la centralidad de cercanía, se interpreta como que tan cercano y central es un vértice con respecto a otros en la red, esto calculando los caminos mas cortos de cada vértice a los otros [4].

Esto quiere decir que los metabolitos resultantes son por los que mas se puede pasar de una reacción a otra, siendo los que mas son producidos y a la vez los que producen reacciones.

Respecto a las reacciones, estas son por las

que mas pasan metabolitos y salen productos, apreciándose que tienen mayor centralidad de cercanía que los metabolitos, lo cual indica que en la red las reacciones son mas centrales y cercanas a los demás vértices que los metabolitos.

- **Centralidad de intermediación:** En metabolitos, los resultados fueron

1. *h_c*, con un grado de 0,6939
2. *h2o_c*, con un grado de 0,3215
3. *h_p*, con un grado de 0,2276

Por otro lado, para las reacciones fueron:

1. *PPA*, con un grado de 0,0253
2. *NO3R1*, con un grado de 0,0216
3. *CYTBDpp*, con un grado de 0,0208

Respecto a la centralidad de intermediación, se puede interpretar como los nodos que se encuentran en mas caminos de un vértices a otro, es decir, los que mas conectan vértices entre sí [1].

Respecto a los metabolitos, los resultantes son los que mas conectan caminos entre los vértices de la red.

Así mismo, en las reacciones estas son las que mas conectan caminos entre vértices de la red, pero se puede evidenciar que tienen un número mucho mas bajo que los metabolitos, por lo cual se puede concluir que los metabolitos son mucho mas importantes a la hora de conectar caminos entre vértices que las reacciones.

- **Centralidad de Katz:** En metabolitos, los resultados fueron

1. *h_c*, con un grado de 0,4068
2. *nadp_c*, con un grado de 0,2668
3. *nad_c*, con un grado de 0,2581

Por otro lado, para las reacciones fueron:

1. *NADH10*, con un grado de 0,0565
2. *NADH17pp*, con un grado de 0,0565
3. *NAHD5*, con un grado de 0,0564

Respecto a la centralidad de Katz, se puede interpretar como la popularidad de un vértice basado en el grado de centralidad de entrada o

salida de sus vecinos, es decir, en que tan populares sean los mismo dentro de la red. [5]. Así, los metabolitos resultantes son los que tienen vecinos mas importantes dentro de la red, siendo los que mejores salidas y entradas tienen dentro de la red.

Así mismo, las reacciones resultantes son las que tienen mejores vecinos en la red, pero vemos que su número es bastante menor que el de los metabolitos, por lo cual se concluye que hay mejores salidas desde un metabolito que desde una reacción

Red debilmente conexas:

- **Centralidad de grado de entrada:** En metabolitos, los resultados fueron

1. *h_c*, con un grado de 0,1719
2. *adp_c*, con un grado de 0,0632
3. *pi_c*, con un grado de 0,06

Por otro lado, para las reacciones fueron:

1. *ASNS*, con un grado de 0,0009
2. *CBPS*, con un grado de 0,0009
3. *CINDO*, con un grado de 0,0009

Vemos que los resultados tanto en metabolitos como en reacciones son bastante similares a la red fuertemente conexas, solo existiendo un cambio en la tercera mejor reacción. Sin embargo, se puede notar que los grados son ligeramente menores, lo cual puede ser debido a que, al existir mas conexiones al ser debilmente conexas, hay muchas mas comparaciones de grados que se pueden hacer, lo cual disminuye un poco el grado de cada vértice.

- **Centralidad de grado de salida:** En metabolitos, los resultados fueron

1. *h_c*, con un grado de 0,1719
2. *adp_c*, con un grado de 0,0632
3. *pi_c*, con un grado de 0,06

Por otro lado, para las reacciones fueron:

1. *ASNS*, con un grado de 0,0009
2. *CBPS*, con un grado de 0,0009
3. *CINDO*, con un grado de 0,0009

De igual manera que en la red fuertemente conexas, los resultados son análogos en grado

de entrada y de salida, manteniéndose la relación de producir y ser producido que se había evidenciado en la anterior red.

- **Centralidad de cercanía:** En metabolitos, los resultados fueron

1. *lys_Lc*, con un grado de 0,1859
2. *ins_c*, con un grado de 0,1859
3. *ptrc_c*, con un grado de 0,1859

Por otro lado, para las reacciones fueron:

1. *ADD*, con un grado de 0,2292
2. *GUAD*, con un grado de 0,2286
3. *CSND*, con un grado de 0,2285

Vemos que los metabolitos y reacciones resultantes son exactamente iguales, con la diferencia de que el número en cada uno disminuye con respecto a la red fuertemente conexa, esto por lo supuesto en las centralidades de grados. Así mismo, se mantiene que las reacciones son mas centrales que los metabolitos en la red.

- **Centralidad de intermedicación:** En metabolitos, los resultados fueron

1. *h_c*, con un grado de 0,4413
2. *h2o_c*, con un grado de 0,2031
3. *h_p*, con un grado de 0,1426

Por otro lado, para las reacciones fueron:

1. *ATPS4rpp*, con un grado de 0,0426
2. *PYK*, con un grado de 0,0354
3. *PPA*, con un grado de 0,0148

Este es el tipo centralidad con la que mas se evidencian cambios, siendo que en las reacciones cambia totalmente el resultado de las mismas, siendo algunos de los números de las reacciones mayores que en la red fuertemente conexa. Sin embargo, se sigue manteniendo que los metabolitos son mas importantes a la hora de conectar caminos entre vértices que las reacciones, por un margen bastante alto, aunque menor que en la red fuertemente conexa.

- **Centralidad de Katz:** En metabolitos, los resultados fueron

1. *h_c*, con un grado de 0,4052
2. *nadp_c*, con un grado de 0,2664

3. *nad_c*, con un grado de 0,2578

Por otro lado, para las reacciones fueron:

1. *NADH17pp*, con un grado de 0,0565
2. *NADH10*, con un grado de 0,0565
3. *NAHD5*, con un grado de 0,0564

Este tipo de centralidad es el que menos varía entre las dos redes, existiendo solo un cambio de orden entre las primeras dos reacciones. Así mismo, los números tanto en metabolitos como en reacciones no varían mucho, siendo una disminución bastante pequeña, concluyendo que la importancia de los vecinos de cada vértice no varía mucho de red a red.

10. Conclusiones finales.

En este proyecto se pudo evidenciar como, por medio de la teoría de grafos y, en específico, por medio del análisis de centralidad de una red conexa, se pueden entender y visualizar mucho mas claramente como funcionan las conexiones y reacciones de una red metabólica, tarea que puede resultar muy difícil de realizar por otros métodos utilizados previamente en diversas areas de la biología, donde es de interés el estudio de este tipo de redes.

Con los resultados obtenidos es posible saber mas acerca del comportamiento tanto de metabolitos como de reacciones en la red, siendo posible con estos datos manipular u optimizar la red según se requiera, esto con la finalidad de realizar procesos como curaciones y llenado de vacios con la finalidad de que la red funcione según requiera el investigador.

A su vez, este proyecto puede ser utilizado como guía para el análisis de otras redes metabólicas mucho mas complejas, ya que como se pudo evidenciar, aunque sean grafos con bastantes vértices, aristas y conexiones, es posible realizar un análisis por centralidad, de modo que se puede obtener los nodos mas relevantes para trabajar con la red de una mejor manera.

Referencias

- [1] C.Perez, R.Germon. "Automating Open Source Intelligence". En: 1 (2016).

- [2] Contreras Gamba, C. “Curación de redes metabólicas con teoría de grafos.” En: (2016). Uniandes, págs. 17-19.
- [3] Contreras Gamba, C. “Curación de redes metabólicas con teoría de grafos.” En: (2016). Uniandes, págs. 24-28.
- [4] D.L.Hansen, B.Shneiderman, M.A.Smith, I.Himmelboim. “Analyzing Social Media Networks with NodeXL”. En: 1 (2020).
- [5] E.Nathan, G.Sanders, J.Fairbanks, V.E.Henson, D.A.Bader. “Procedia Computer Science”. En: 108 (2017), 68-78.
- [6] J.Goldbeck. “Introduction to social media investigation: A Hands-On Approach”. En: 1 (2015).
- [7] McCloskey, Douglas, Palsson, Bernhard Ø, Feisty Adam M. “Basic and applied uses of genome- scale metabolic network reconstructions of Escherichia coli”. En: (2013).
- [8] Sociedad Española de Bioquímica y Biología Molecular. “La metabolómica: un déjà vu por la historia de la bioquímica.” En: *SEBBM* (2015), pág. 185.
- [9] Valerie P. Cortés. “MetaPenta: visualización y análisis de redes metabólicas”. En: 1 (2021), 1-3.
- [10] Wasserman Faust. “Centralidad y prestigio”. En: (2013), 191-240.