

Trabajo Computacional 2

Estudio de la regla de la centralidad-letalidad en cuatro redes de interacción de proteínas de *Saccharomyces cerevisiae*

Heli Magali García Álvarez, Juan Ignacio Gossn, Santiago Scheiner

Resumen

En este trabajo se reprodujeron algunos de los procedimientos presentados en el trabajo de Zotenko *et al.* [1] para redes de interacciones de proteínas.

El eje principal del trabajo fue el análisis de la regla de la centralidad-letalidad, que establece que nodos de alto grado en una red de proteínas tienden a corresponderse con aquellas proteínas que resultan esenciales para la red, o lo que es lo mismo, con aquellas proteínas letales si se deletan sus marcos de lectura abiertos correspondientes en el organismo modelo de estudio. El objetivo fue entonces el estudio de cuatro redes de proteínas de la levadura *Saccharomyces cerevisiae*, a fin de obtener una explicación biológica para la relación entre la esencialidad de las proteínas y las propiedades topológicas de las redes.

Los resultados mostraron ser consistentes con lo obtenido en el trabajo de Zotenko *et al.*, y siguiendo la línea de este trabajo, se desestimó la hipótesis de Jeong *et al.* [5], que sostiene que los hubs esenciales son aquellos responsables de mantener la conectividad de la red, y la hipótesis de He *et al.* [3], que afirma que la mayoría de las proteínas son esenciales porque participan en una o más interacciones esenciales distribuidas uniformemente al azar lo largo de la red.

1. Introducción

Los conceptos de centralidad y letalidad suelen relacionarse con frecuencia argumentando que los nodos de alto grado en una red de interacciones de proteínas tienden a corresponderse con las proteínas que son esenciales para el funcionamiento de la red y por ende, para el proteoma del organismo. Esta relación sugiere entonces que la prominencia topológica de una proteína en la red es un parámetro útil para predecir su esencialidad.

Frente a estas hipótesis, en el trabajo de Zotenko *et al.* [1] se demostró que estas suposiciones carecen de fundamentos basados en las propiedades de las redes de proteínas.

Este trabajo tuvo como objetivo la extensión del procedimiento llevado a cabo por Zotenko *et al.* a cuatro redes de interacción de proteínas correspondientes a *Saccharomyces cerevisiae*. En concreto, se analizaron diferentes métodos para determinar la centralidad de un nodo de la red y se la contrastó con los resultados del trabajo de He *et al.* [3]. En su trabajo, He desafía la hipótesis de que la esencialidad de un nodo sea una característica global de la red y propone que la esencialidad de una proteína tiene origen en el hecho de estar involucrada en al menos un enlace esencial. Pensado de esta manera, nodos de alto grado tienden a ser esenciales por estar involucrados en más enlaces, lo que aumenta la probabilidad de que uno de ellos sea esencial.

Las dos primeras redes analizadas en este trabajo corresponden a los métodos Y2H (técnica del doble híbrido en levadura) y AP-MS (cromatografía de afinidad seguido de espectrometría de masas). Las otras dos corresponden a redes de interacción provenientes de la literatura (redes de interacción de proteínas formada a partir de la aparición simultánea en artículos científicos).

El eje principal de este trabajo fue entonces el análisis de la relación entre la centralidad de los nodos de estas redes y la esencialidad de las proteínas a las que estos nodos representan.

El concepto de *centralidad* consiste en la determinar qué nodos resultan más importantes para la red. Existen diferentes maneras de establecer qué nodos son importantes en una red, y la utilización de cada una dependerá del problema que se esté modelando [2].

La manera más simple de establecer esta magnitud es la *centralidad de grado*, que consiste en determinar la centralidad en la red a partir del grado de los nodos. De esta manera, un nodo resulta relevante cuando posee muchas conexiones con otros nodos de la red.

Una extensión natural de la centralidad de grado es la *centralidad de autovalor*. En este caso, la importancia de los nodos está dada por sus vecinos, pero teniendo en cuenta a su vez la importancia de cada uno de estos. De este modo, no todos los vecinos de un nodo son equivalentes, y la importancia de un nodo aumenta si tiene conexiones con nodos que son importantes en sí mismos. Por lo tanto, la centralidad de autovalor proporciona a un nodo una importancia que equivale a la suma de las importancias de sus vecinos. La idea es entonces hacer una estimación inicial para la centralidad x_i de cada nodo i , como puede ser, por ejemplo: $x_i = 1$. A partir de estos valores iniciales es posible entonces realizar estimaciones de manera iterativa de modo de asignarle a la centralidad x_i , la suma de las centralidades de sus vecinos del nodo i :

$$x' = \sum_j A_{ij} x_j$$

donde A_{ij} es un elemento de la matriz de adyacencia. Escrito de manera matricial, se tiene: $x' = Ax$, lo que significa que x converge a un autovalor de la matriz de adyacencia.

Una manera completamente diferente de medir la centralidad de los nodos es la *centralidad de cercanía*, que mide la distancia que lo separa de los otros nodos de la red. Si d_{ij} es la distancia mínima desde el nodo i hasta el nodo j , la media de las distancias mínimas desde el nodo i hasta los demás nodos es:

$$l_i = \frac{1}{n} \sum d_{ij}$$

Entonces, la centralidad de cercanía puede calcularse como:

$$C_i = \frac{1}{l_i}$$

Finalmente, otra forma de medir la centralidad, que tampoco está relacionada con las anteriores es la *centralidad de intermediación* de un nodo, que consiste establecer cuántos caminos mínimos entre otros dos nodos pasan por el nodo que se analiza. En este sentido, un nodo se vuelve relevante cuando comunica muchos otros nodos entre sí, tomando el camino óptimo entre ellos.

Los criterios de centralidad detallados arriba fueron los utilizados en este trabajo.

2. Resultados y discusión

Como se mencionó antes, los casos estudiados consistieron en cuatro redes de interacción de proteínas correspondientes a la levadura *Saccharomyces cerevisiae*.

En primer lugar, se estudiaron dos redes de interacciones relevadas mediante los métodos experimentales de Y2H (técnica del doble híbrido en levadura) y AP-MS (cromatografía de afinidad seguido de espectrometría de masas).

Por otro lado, se analizaron también dos redes de interacción provenientes de la literatura, a las que se referirá en este trabajo como LIT (red de interacción de proteínas formada a partir

de la búsqueda en artículos científicos) y LIT_{Reguly} (red formada a partir de la búsqueda en artículos científicos, estructurada por Reguly *et al.* [4]).

2.1. Características de las redes analizadas

En el cuadro 1 se muestran la cantidad de nodos y enlaces, el grado medio y el coeficiente de *clustering* medio de cada red.

| | Nodos | Enlaces | Grado medio | C_{avg} |
|----------------|-------|---------|-------------|-----------|
| Y2H | 2017 | 2929 | 2.904 | 0.046 |
| AP-MS | 1622 | 9069 | 11.182 | 0.554 |
| LIT | 1535 | 2924 | 3.810 | 0.293 |
| LIT_{Reguly} | 3307 | 11858 | 7.171 | 0.261 |

Cuadro 1: Propiedades estructurales de las redes de interacción de proteínas testeadas.

En primer lugar, se observa que se trata de cuatro redes de dimensiones comparables, en el sentido de que el número de nodos presente en cada una difiere como mucho en un factor dos.

Por otro lado, puede verse que tanto el número de enlaces como el grado medio y el coeficiente de clustering medio de la red Y2H son notoriamente menores que para la red AP-MS. Esto se condice con lo que se espera si se tiene en cuenta los procedimientos de relevo de interacciones para cada red. Debido a la naturaleza de la red Y2H, para la que se relevan interacciones binarias en un ensayo de naturaleza muy controlada, no es sorprendente que los coeficientes calculados resulten menores que para la red AP-MS, para la que se relevan complejos de proteínas interactuantes relacionados a una proteína de interés y por lo tanto, tiende a sobreestimar el número de interacciones. Nuevamente, el alto coeficiente de clustering medio para la red AP-MS, el más alto entre todas las redes estudiadas, puede considerarse un indicio del sesgo que se introduce al relevar y anotar las interacciones proteicas mediante esta técnica experimental.

En cuanto a las redes que surgen del relevo de datos presentes en la literatura científica, se observa que la red LIT_{Reguly} [4] posee alrededor del doble de nodos que la red LIT y más de tres veces el número de enlaces que la última. Esto es esperable ya que la red LIT_{Reguly} fue construida a partir de una búsqueda más exhaustiva de interacciones en la literatura científica, en la cual se curaron manualmente más de 31,793 *abstracts* y publicaciones online [4]. Al aproximadamente duplicarse el número de nodos y triplicarse el número de enlaces con respecto a la red LIT, es lógico encontrar un grado medio de casi el doble para LIT_{Reguly} en comparación con LIT. El coeficiente de clustering medio es similar para ambas redes, lo cual muestra que a pesar de que la red LIT es más pequeña porque contiene menos nodos y enlaces, los nodos de la misma poseen vecinos que están conectados entre sí en promedio de la misma forma que en la red LIT_{Reguly} .

Tener en cuenta las características de las redes se torna importante en los análisis siguientes, debido a que varios procedimientos deben ser relajados en algunos casos puntuales, como se detalla más adelante. En particular, debido a que la red Y2H es menos densa (en términos de enlaces) y menos conexas que el resto de las redes analizadas, lo cual se evidencia con el grado medio y el coeficiente de clustering medio, se contabilizarán pares de proteínas no adyacentes con 1 vecino común o más, en vez de aquellos pares con 3 vecinos o más en el procedimiento correspondiente al cuadro 4.

Con la finalidad de esbozar una idea acerca de la semejanza entre las redes estudiadas, en el cuadro 2 se muestran los grados de superposición de los enlaces presentes en cada red. Los grados calculados corresponden a la proporción de enlaces presentes en una red, que se encuentran a su vez en otra de las redes.

| | Y2H | AP-MS | LIT | LIT _{Reguly} |
|-----------------------|-------|-------|-------|-----------------------|
| Y2H | 1.000 | 0.089 | 0.088 | 0.164 |
| AP-MS | 0.029 | 1.000 | 0.143 | 0.278 |
| LIT | 0.089 | 0.443 | 1.000 | 0.978 |
| LIT _{Reguly} | 0.040 | 0.212 | 0.241 | 1.000 |

Cuadro 2: Grado de superposición de enlaces entre las redes de proteínas testeadas.

Es importante notar que para cada par de redes se obtienen dos valores de grado de superposición diferentes, dado que la intersección o número compartido de enlaces entre las mismas es uno, pero dicho valor representa una porción diferente del total de enlaces para cada red del par considerado. La manera de interpretar los valores de la tabla es entonces la siguiente: cada fila corresponde a una red, y cada valor de esa fila es la fracción de nodos de esa red que se encuentran presentes también en cada una de las otras redes estudiadas. Por ejemplo, 8.9 % de los nodos presentes en la red Y2H se encuentran también en la red AP-MS.

Por otra parte, en la figura 1 se muestra qué fracción de los nodos con grado mayor o igual a un umbral definido (aquellos que cumplen con esta condición son *hubs*) son esenciales. El eje horizontal muestra la fracción de nodos del total de la red que fueron designados como hubs.

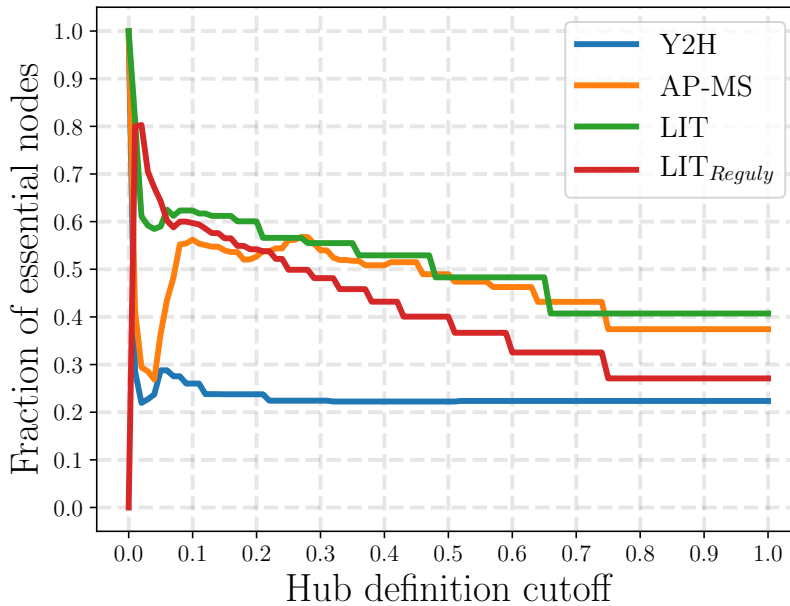


Figura 1: Relación entre el grado y la esencialidad de los nodos en las redes analizadas.

Para entender un poco más lo que muestra este gráfico, consideremos los valores extremos de 0 y 1 en el eje horizontal. Cuando todos los nodos de la red son considerados como hubs (es decir que el eje de las abscisas es igual a 1), la fracción de nodos esenciales en el eje de las ordenadas es igual a la proporción de nodos esenciales del total de nodos de la red. Lo contrario sucede cuando el eje x es igual a 0: al no considerar a ningún nodo como hub, la fracción de nodos esenciales del total de hubs es por definición 0. En este gráfico se observa una clara tendencia, a medida que se relaja el umbral de designación de hubs, la fracción de nodos esenciales dentro de los hubs considerados baja. Esto implica que los nodos con alto grado k o hubs están enriquecidos en nodos esenciales. Cabe destacar que se observan algunas fluctuaciones en las curvas entre los valores de 0 y 0.1 del eje x (especialmente para la red AP-MS, en la cual hay una baja abrupta en la cantidad de nodos esenciales en el intervalo

mencionado), lo cual indica que el enriquecimiento sostenido en hubs esenciales a medida que aumenta el grado k umbral de designación de un hub se ve interrumpido para el 10 % de nodos con k mayor. En relación a la red Y2H, a diferencia de las otras 3 redes analizadas, la tendencia mencionada es mucho más débil.

La figura 1 reproduce los resultados pioneros de Jeong *et al.* [5], luego también confirmados en estudios posteriores, que muestran que los hubs en las redes de proteínas de *Saccharomyces cerevisiae* contienen más nodos esenciales que lo que se esperaría por azar. Jeong *et al.* sostienen que este fenómeno puede ser atribuido al rol central que juegan los hubs en mediar interacciones entre grupos numerosos de proteínas poco conectados entre sí.

2.2. Análisis de vulnerabilidad

Una manera de medir el impacto de la remoción de nodos sobre la conectividad de una red es monitorear la disminución en tamaño de la componente gigante. La figura 2 muestra la vulnerabilidad de las redes estudiadas frente a la remoción de las proteínas más centrales.

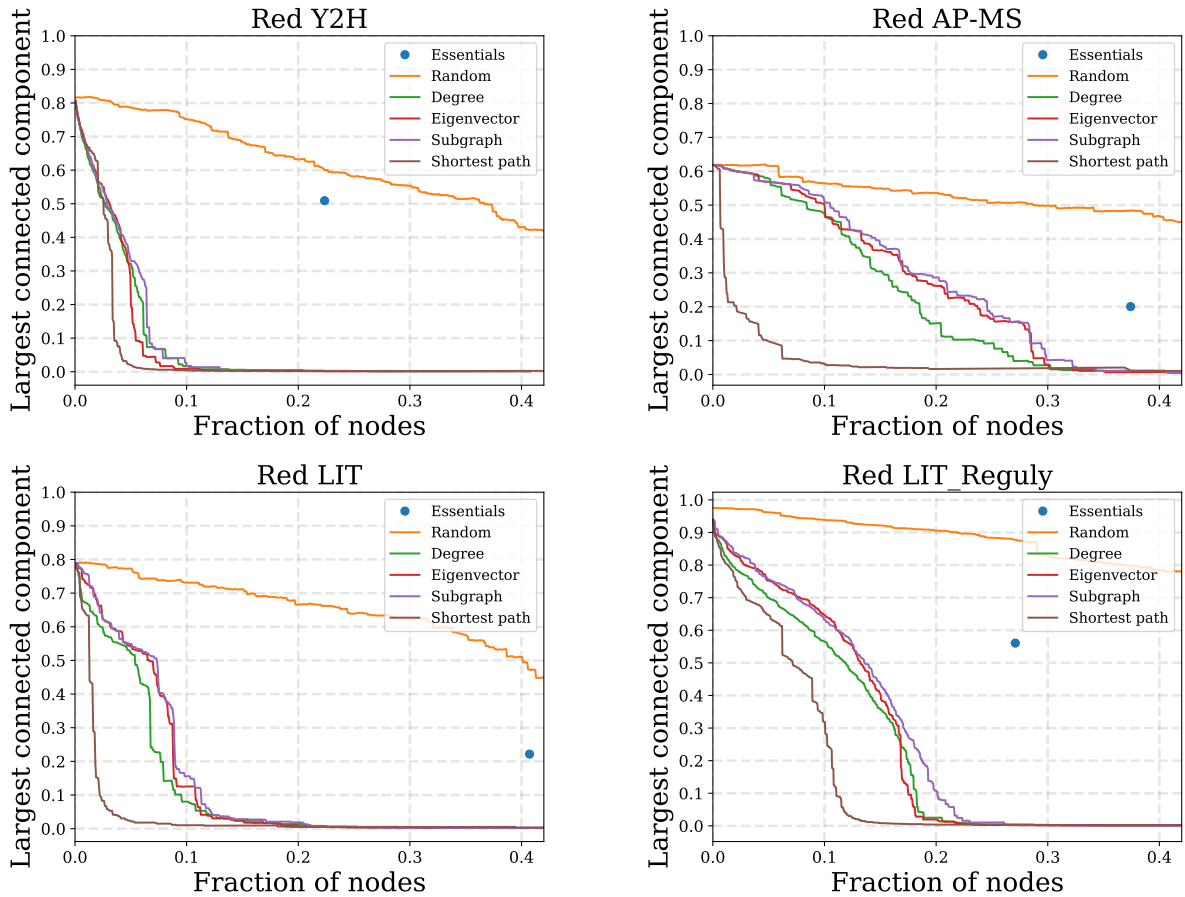


Figura 2: Vulnerabilidad frente al ataque contra las proteínas mas centrales de la red.

Una vez más, el criterio utilizado para cuantificar el impacto de este proceso sobre la red es el registro de la cantidad de nodos presentes en la componente más grande de la red a medida que se quitan los nodos de a uno a la vez.

La eliminación de los nodos se realiza siempre sobre el nodo restante de máxima centralidad. Este proceso se repite para diferentes criterios de centralidad: de grado, de autovector, de subgrafos y de intermediación de caminos mínimos. La norma adoptada para la remoción sucesiva de nodos es el cálculo de la centralidad en cada paso. De esta manera, la centralidad de los nodos se renueva con cada paso y se elimina la posibilidad de incertezas debidas al uso de valores de centralidad sin actualizar.

Esto se compara además con la remoción de nodos al azar. Finalmente, se comparan las curvas obtenidas con lo que se obtiene al remover todos los nodos esenciales en un único paso.

Puede observarse a simple vista en las figuras de las cuatro redes que para mantener la conectividad de la red, los nodos esenciales no resultan de mayor importancia que los no esenciales. Esto se ve en el hecho de que la remoción de los nodos esenciales queda muy por encima de las curvas trazadas por la remoción según los métodos de centralidad. Esto significa que la eliminación de nodos según su centralidad (cualquiera sea la definición de centralidad que se utilice) consigue “desarmar” la red mucho más rápidamente que si se eliminan todos los nodos esenciales.

Estos resultados, como los mostrados en la figura 3 de Zotenko *et al.* [1], desestiman la hipótesis de Jeong *et al.* [5] quienes explican el enriquecimiento de los hubs en proteínas esenciales atribuyéndole a dichos hubs esenciales la propiedad topológica de sostener la conectividad de la red.

Para analizar aún más las causas de este resultado, se estudió también el impacto de la eliminación de nodos esenciales comparado con la remoción de un número equivalente de nodos no esenciales tomados al azar con la misma distribución de grado. Los resultados de este análisis se muestran el cuadro 3.

| | Esencial | No esencial al azar |
|-----------------------|----------|---------------------|
| Y2H | 0.656 | 0.599 ± 0.021 |
| AP-MS | 0.558 | 0.421 ± 0.027 |
| LIT | 0.657 | 0.582 ± 0.010 |
| LIT _{Reguly} | 0.929 | 0.765 ± 0.006 |

Cuadro 3: Impacto de la remoción de las proteínas esenciales comparado con la remoción de un número equivalente de proteínas no esenciales tomados al azar con la misma distribución de grado.

Debido a la imposibilidad de quitar un conjunto equivalente de proteínas no esenciales con exactamente los mismos grados que el conjunto de proteínas esenciales removido, se apeló a un criterio más flexible para poder llevar a cabo el proceso de remoción de proteínas de la red. En primer lugar, se dividieron a las proteínas esenciales y no esenciales (por separado) en deciles según sus grados, teniendo en cuenta la distribución de grados de la red original. Se contó la cantidad de proteínas esenciales en cada decil y se tomó una muestra al azar de proteínas no esenciales del mismo tamaño en el decil correspondiente. Al finalizar el muestreo, se removieron las proteínas esenciales y las no esenciales elegidas al azar y se calculó la fracción de nodos en la componente gigante. El proceso de remoción al azar de proteínas no esenciales se repitió 30 veces y se calculó el promedio y el desvío estándar de la fracción de nodos en la componente gigante luego de la remoción. Los resultados obtenidos no concuerdan, una vez más, con la hipótesis de Jeong *et al.* [5], ya que la remoción de nodos no esenciales tomados al azar tiene un efecto más disruptivo sobre la fracción de nodos remanente en la componente gigante que la remoción de los nodos esenciales, para todas las redes estudiadas. Este resultado reafirma la idea de que los nodos esenciales carecen de importancia de origen topológico. De todos modos, se esperaba que no hubiese una diferencia significativa entre los valores correspondientes a los nodos esenciales y a los no esenciales, tal como lo reportado en la tabla 3 de Zotenko *et al.* [1]. Esta diferencia encontrada puede adjudicarse al modo en que se flexibilizó el criterio de remoción de nodos, en particular, a los casos en que por falta de nodos no esenciales de igual grado que el de la proteína esencial, debió eliminarse nodos de mayor grado.

2.3. Estudio del rol de las interacciones esenciales

Para estudiar la hipótesis proveniente del trabajo de He *et al.* [3], quienes afirman que algunas proteínas son esenciales porque participan en interacciones esenciales distribuidas uniformemente en la red, es necesario estimar la probabilidad de que un nodo sea esencial. En primer lugar, se asigna α a la probabilidad de que un nodo sea esencial en una determinada red debido a que participa en un enlace esencial, y β a la probabilidad de que un nodo resulte esencial por algún otro motivo. La probabilidad P_E de que una proteína de grado k sea esencial es entonces:

$$P_E = 1 - (1 - \beta)(1 - \alpha)^k$$

Tomando el logaritmo de ambos miembros, se obtiene:

$$\log(1 - P_E) = k \log(1 - \alpha) + \log(1 - \beta)$$

A partir de esta última expresión es posible realizar un ajuste lineal teniendo en cuenta los datos correspondientes a la frecuencia de nodos esenciales presentes en la red en función de su grado k , y de esta manera obtener los coeficientes α y β .

En la figura 3 se observan estos ajustes para cada red, junto con los valores obtenidos en cada uno.

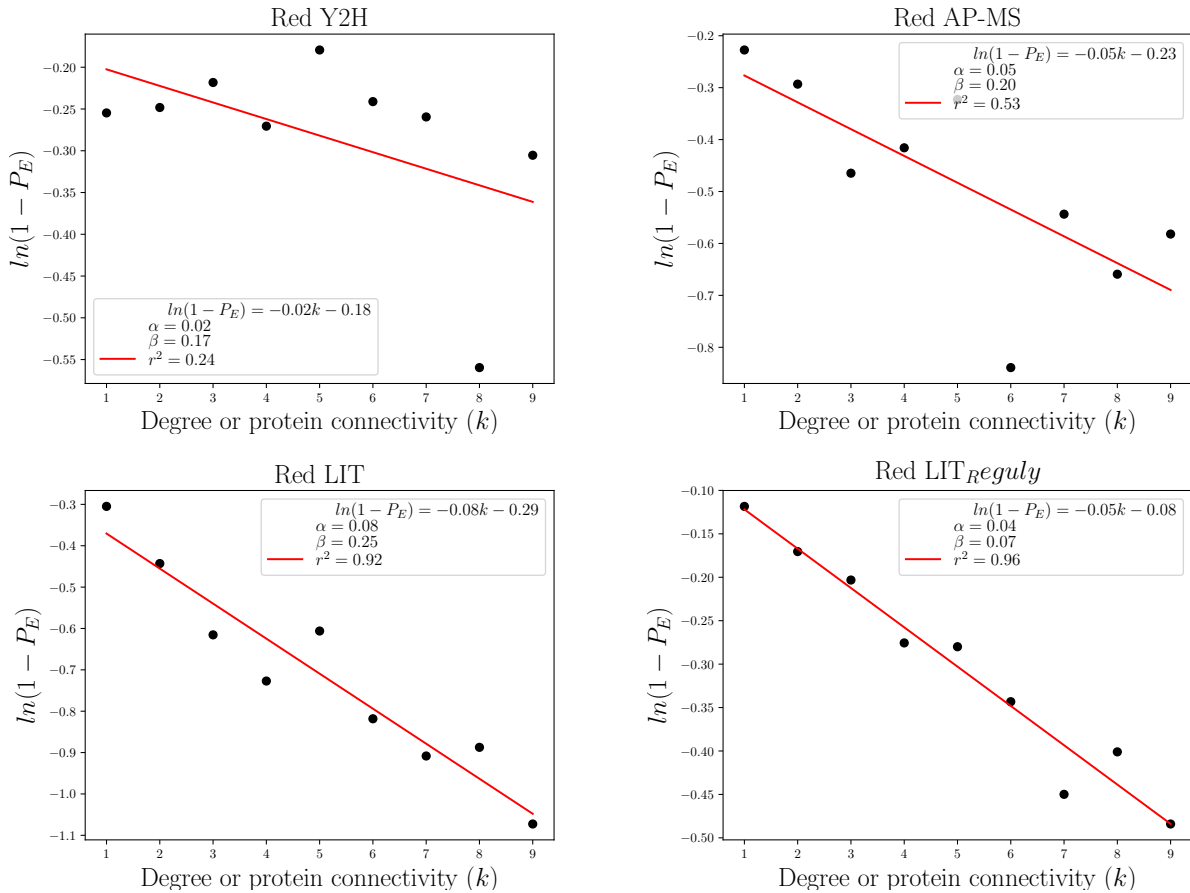


Figura 3: Relación entre la probabilidad de que una proteína sea esencial (PE) y la conectividad o el grado k de la proteína.

Con los valores de los coeficientes α y β , es posible entonces calcular algunas propiedades que debería cumplir la red. En particular, puede estimarse el número de pares de nodos no

adyacentes de igual tipo (con vecinos en común) que deberían encontrarse presentes en la red y contrastar estos valores con los valores reales de la red.

Para construir el cuadro 4, primero se calcula el número de pares de nodos no adyacentes totales con 3 o más vecinos en común, para el caso de las redes AP-MS, LIT, y LIT_{Reguly}, y con 1 o más vecinos en común para el caso de la red Y2H, por las razones mencionadas en la sección 2.1. Luego, de estos pares de nodos no adyacentes relevados, se registra la cantidad de pares constituidos por proteínas del mismo tipo, es decir ambas esenciales o ambas no esenciales. Por último, se estima la cantidad esperada de pares con nodos del mismo tipo teniendo en cuenta el modelo de interacciones esenciales de He *et al.*. Para cada par de nodos no adyacentes de una dada red, se calcula la probabilidad del evento nodo esencial P_E y nodo no esencial $1 - P_E$ para cada proteína del par, teniendo en cuenta su grado k . Tanto para obtener la probabilidad de par de nodos esenciales, como la de par de nodos no esenciales, se multiplican las probabilidades de los eventos ya calculados, ya se consideran independientes. La probabilidad de par de nodos del mismo tipo para un dado par de nodos no adyacentes resulta ser la siguiente:

$$P_{\text{nodo}_{ij} \text{ mismo tipo}} = 1 - P_{E_i} - P_{E_j} + 2P_{E_i}P_{E_j}$$

La sumatoria de $P_{\text{nodo}_{ij} \text{ mismo tipo}}$ sobre todos los pares de nodos ij será igual al número esperado de nodos del mismo tipo.

Los resultados de este análisis pueden observarse en el cuadro 4.

| | Cantidad total de pares | Pares de igual tipo | Cantidad esperada de pares de igual tipo |
|-----------------------|-------------------------|---------------------|--|
| Y2H | 23070 | 15085 | 14075 |
| AP-MS | 11614 | 5908 | 7758 |
| LIT | 730 | 389.0 | 397 |
| LIT _{Reguly} | 10777 | 6187 | 5634 |

Cuadro 4: Diferencia entre la cantidad observada y esperada de número de pares de proteínas del mismo tipo (ambas son esenciales o no esenciales).

Como se muestra en el cuadro 4, hay diferencias apreciables entre la cantidad de pares del mismo tipo y la cantidad de pares esperados del mismo tipo calculados bajo el modelo de He *et al.*, especialmente para la red AP-MS. Si estas diferencias fuesen estadísticamente significativas, rechazaríamos el modelo de las interacciones esenciales, tal como muestra la tabla 5 del trabajo de Zotenko *et al.*

3. Conclusiones

En primer lugar, como se mencionó en la sección anterior, los resultados presentados en el cuadro 1 son consistentes con lo esperado para las redes analizadas si se tienen en cuenta los métodos utilizados para el relevo de las interacciones entre proteínas. Tener en cuenta las características topológicas básicas de las redes nos permitió adaptar los métodos para análisis posteriores, especialmente en lo que refiere al cuadro 4.

Comenzamos nuestro análisis estudiando la relación entre el grado de un nodo y la esencialidad de la proteína que representa el mismo en la red. Hallamos que existe una correlación entre ambas variables: nodos de grados más altos o hubs están enriquecidos en proteínas esenciales. Cabe destacar que esta correlación fue mucho más débil para el caso de la red Y2H, que es menos densa en enlaces y conexa que las otras redes analizadas.

Debido a que recopilamos los resultados pioneros de Jeong *et al.* [5], quienes indican que hay una sobrerrepresentación de proteínas esenciales en nodos de alto grado o hubs, nos focalizamos en estudiar la regla de la centralidad-letalidad planteada por el mismo autor. Es

así que realizamos, en la sección 2.2, un análisis de vulnerabilidad frente a la remoción de proteínas o nodos esenciales y no esenciales de las redes. Tal como se observa en la figura 2, la remoción de los nodos esenciales queda muy por encima de las curvas trazadas para la remoción según los diferentes métodos de centralidad empleados. Esto significa que la misma es menos disruptiva para la conectividad global de la red medida como la fracción de nodos remanente en la componente gigante. En suma, los nodos esenciales no juegan un rol decisivo para sostener la conectividad de la red.

Debido a que descartamos la hipótesis de Jeong *et al.*, seguimos en la búsqueda de algún determinante topológico de la red correlacionado con la esencialidad de los nodos. En la sección 2.3 exploramos la hipótesis de He *et al.* [3], que se centra en las interacciones esenciales distribuidas uniformemente al azar en la red. Para estudiar lo propuesto por este autor, planteamos su modelo y estimamos así la probabilidad de que un nodo de la red corresponda a una proteína esencial, siendo que este modelo de interacción de proteínas esenciales asume que si dos proteínas no interactúan, la esencialidad de una no se ve afectada por la esencialidad de la otra en un mismo par. En particular, se estudió el caso de nodos no adyacentes con vecinos en común en las redes, para los cuales también debería valer el supuesto de la independencia de eventos mencionado anteriormente. Se comparó el número esperado de pares de proteínas del mismo tipo con el número real de pares de proteínas del mismo tipo relevados de las redes, tal como se muestra en la figura 4, obteniendo una diferencia apreciable entre ambos valores. Debido a que no tenemos estadística de los datos no podemos rechazar la hipótesis de He *et al.*. No obstante, podemos afirmar que nuestros resultados están en concordancia con lo reportado por Zotenko *et al.*

Para finalizar, sería importante mencionar que una posible continuación de este trabajo sería testear con las redes de proteínas estudiadas la hipótesis alternativa planteada por Zotenko *et al.*, que sostiene que la mayoría de los hubs son esenciales debido a su participación en ECO-BIMs (por sus siglas en inglés, Essential Complex Biological Modules), un grupo de proteínas densamente conectadas con funcionalidad biológica compartida, las cuales están enriquecidas en nodos esenciales.

Referencias

- [1] ZOTENKO, Elena, *et al.* (2008). "Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality." PLoS Computational Biology, vol. 4, no 8, p. e1000140.
- [2] NEWMAN, Mark (2010). Networks. Oxford University Press.
- [3] HE, Xionglei; ZHANG, Jianzhi (2006). "Why do hubs tend to be essential in protein networks?." PLoS Genetics, vol. 2, no 6, p. e88.
- [4] REGULY, Teresa, *et al.* (2006). "Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*." Journal of Biology, vol. 5, no 4, p. 11.
- [5] JEONG, Hawoong, *et al.* (2001). "Lethality and centrality in protein networks." Nature, vol. 411, no 6833, p. 41.