

Análisis crítico de dos modelos de explicación para la regla de centralidad-letalidad en redes proteicas de *Saccharomyces cerevisiae*

Emanuel Chironi, Federico Seivler, Lucas Alonso, and Francisco Correa
Universidad de Buenos Aires
Redes complejas
Trabajo Computacional N 2

La regla de centralidad y letalidad es un ejemplo de relación entre funcionalidad biológica y topología de redes, la cual ha sido reportada por la literatura especializada. Sin embargo, el origen de tal comportamiento permanece elusivo. Al respecto, se han propuesto distintas hipótesis para intentar explicar la regla. En este trabajo se revisan críticamente dos de esos modelos, a partir de la reconstrucción de los pasos dados por Zotenko y sus colaboradores, en la revisión de ambas hipótesis. Para eso se utilizaron cuatro redes de interacción de proteínas de la levadura *Saccharomyces cerevisiae*. Se testearon ambas hipótesis a partir de los datos extraídos de las cuatro redes mencionadas y se obtuvieron resultados parcialmente análogos a los reportados por Zotenko y colaboradores. En consecuencia, se concluyó que la primera hipótesis analizada debería descartarse, pero que la evidencia no era suficiente para descartar la segunda.

I. INTRODUCCIÓN

El análisis de la relación entre las funciones de proteínas y otras macromoléculas con la estructura topológica de las redes de interacción en las que están inmersas, es un campo importante de estudio en biología de sistemas. Un ejemplo del tipo de relaciones que se estudian es la llamada Regla de Centralidad y Letalidad. Esta regla indica que hay una correlación positiva entre la esencialidad biológica de ciertas proteínas y la posición que éstas ocupan en la red de interacciones. Concretamente, una proteína de alto grado (hub) tiene tres veces más chances de ser esencial que una que no lo es. Esta regla fue postulada por primera vez por Jeong y colaboradores.

Sin embargo, las razones por las cuáles la regla existe son más difíciles de determinar y se han planteado varias hipótesis que intentan dar cuenta de ella. Dos hipótesis particularmente relevantes fueron las formuladas por el propio Jeong y colaboradores, y también por He y colaboradores. La pregunta causal podría enunciarse como: Qué hace esenciales a las proteínas y qué relación tiene esa esencialidad con el grado que detentan en la red?

La hipótesis de Jeong es básicamente que las proteínas son esenciales porque mantienen conectada a la red. Pero dado que la centralidad de grado mide el rol de un nodo en la conectividad de la red, se deduce que las proteínas esenciales tenderán a tener mayor grado.

La hipótesis de He es que las proteínas son esenciales porque participan en interacciones esenciales. Dado que estas interacciones están distribuidas uniformemente al azar entre los enlaces de la red, la probabilidad de que una proteína se involucre en enlaces esenciales aumenta con la cantidad de enlaces que tiene con otras proteínas, es decir, con su grado.

Ambos grupos aportaron evidencia a favor de sus respectivas explicaciones. En particular, He aportó no solamente evidencia a favor de su explicación, sino que

procuró aportar evidencia en contra de la hipótesis de Jeong. A su vez, Zotenko y colaboradores analizaron críticamente ambas explicaciones y suministraron evidencia contraria a ambas. En contrapartida, ofrecieron y procuraron sustentar una explicación alternativa a las dos anteriores.

El objetivo de este trabajo computacional consiste en realizar un análisis crítico de las dos hipótesis antes mencionadas, tomando como guía el trabajo de Zotenko, a fin de evaluar la evidencia a favor y en contra de ambas hipótesis. Sin embargo, no entraremos en la explicación alternativa propuesta por el grupo de Zotenko.

II. CARACTERÍSTICAS DE LAS REDES ANALIZADAS

A efectos de reconstruir el análisis hecho por Zotenko consideraremos cuatro redes proteicas: Y2H, LIT, LIT_Reguly y AP-MS. El uso simultáneo de varias redes construidas de manera diferente es relevante para limitar el impacto de sesgos propios del proceso de construcción. Zotenko y colaboradores hacen uso de varias redes porque la literatura anterior había sugerido que incluso la regla de centralidad y letalidad podía ser un efecto espurio. Según esta postura, la correlación entre esencialidad y grado podía deberse al simple hecho de que, dado que las proteínas esenciales han sido más estudiadas, conocemos mejor sus interacciones que las de otras proteínas. Como resultado de ese mayor conocimiento, es esperable que aparezcan más enlaces de proteínas esenciales que de no esenciales, independientemente de la cantidad real de interacciones que las proteínas tienen entre sí.

Antes de avanzar en la contrastación de los modelos propuestos por He y Jeong, es interesante caracterizar las propiedades estructurales de las redes, al menos por razones de completitud, ya que estas propiedades entran en juego en los distintos modelos de explicación. En particular, los parámetros de tamaño pueden ser importantes

a la hora de evaluar el posible sesgo que cada una de las redes pueda tener. La tabla II resume las propiedades para cada una de las redes analizadas.

	# de nodos	# de enlaces	Grado medio	Clustering medio	Clustering global
AP-MS	1622	9070	11.18	0.55	0.62
LIT	1536	2925	3.81	0.29	0.35
LIT_Reguly	3307	11858	7.17	0.26	0.12
Y2H	2018	2930	2.90	0.046	0.024

TABLA I. Resumen de las propiedades estructurales de las cuatro redes analizadas en este trabajo.

Por otra parte, dado que las redes se han construido con diferentes procedimientos, es relevante estudiar cuan parecida es la información que contienen. Es decir, es conveniente estudiar qué proporción de las interacciones reportadas en una red aparecen en las demás. Para esto, medimos el grado de overlap entre las redes. El overlap de una red A respecto de otras redes se define como la fracción de enlaces de A que está presente en las demás. Si bien la cantidad de enlaces compartidos entre dos redes es simétrica, al dividir por la cantidad de enlaces para obtener la fracción, esa simetría se rompe, de modo que la matriz resultante no es simétrica. La tabla II resume el overlap entre las distintas redes. Se puede ver que la red LIT está prácticamente contenida en la red LIT_Reguly, de modo que la segunda es casi una extensión de la primera. Por otra parte, la red Y2H es la que menos información comparte con las demás.

AP-MS	0.14	0.28	0.029
0.44	LIT	0.99	0.089
0.22	0.27	LIT_Reguly	0.042
0.089	0.089	0.17	Y2H

TABLA II. Distribución de overlap de los enlaces entre cada una de las cuatro redes consideradas.

Finalmente, la existencia de la regla de centralidad y letalidad está asociada con el carácter libre de escala de la distribución de grados en la red de proteínas. Este resultado fue específicamente reportado tanto por Jeong como por Zotenko, e incluso confirmado por otros estudios independientes. Dado que la regla es la base del análisis de los modelos que pretenden explicarla, es relevante establecer que esta correlación se mantenga en las cuatro redes que aquí se estudian. Para hacerlo, seguiremos el método utilizado por Zotenko, el cual consiste en graficar la fracción de nodos esenciales dentro de un conjunto de hubs, en función de la fracción total de nodos que se consideran hubs. Para esto se definen los hubs como aquellos nodos que tienen un grado igual o mayor que un grado dado. Este análisis se realizó para las cuatro redes y el resultado se muestra en la Fig. II.

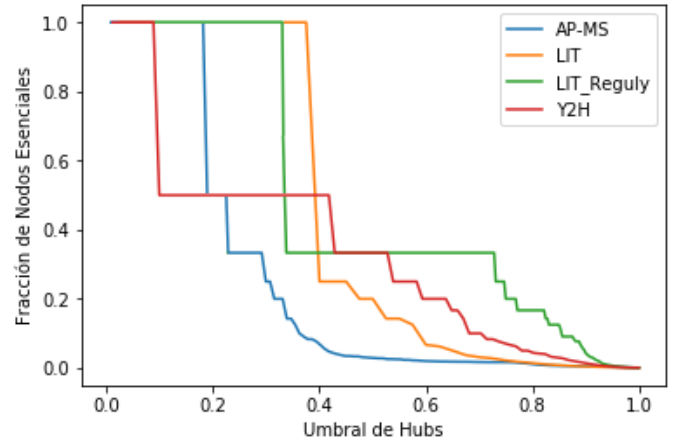


FIG. 1. Fracción de nodos esenciales sobre el total de hubs, en función de la fracción de hubs sobre el total de nodos, para distintos umbrales de definición de hubs.

Como se puede apreciar, la proporción de nodos esenciales entre los hubs se incrementa a medida que el umbral de definición de los hubs se hace más restrictivo. Esto indica que, en proporción, hay más nodos esenciales cuanto mayor es el grado de los nodos considerados. Se observa el cumplimiento estricto de la regla de centralidad y letalidad. Zotenko complementa este análisis cualitativo con la determinación de dos coeficientes de correlación, pero aquí omitiremos este paso.

Dado que la regla de centralidad y letalidad se cumple en todas las redes analizadas, es posible emplearlas para analizar con mayor detalle los modelos explicativos propuestos por Jeong y He.

III. ANÁLISIS DE VULNERABILIDAD

Tal como se adelantó en la introducción, la hipótesis de Jeong es que la esencialidad de las proteínas se debe a que mantienen la conectividad de la red, que a su vez está asociada con el grado de los nodos.

Esta hipótesis tiene dos implicancias cuya ocurrencia puede ser testada. Si la esencialidad de las proteínas se debe a su rol como garantes de la conectividad global de la red, entonces, la remoción de aquellos nodos que tienen una mayor centralidad medida en términos globales debería ser la forma más eficaz de desarmar la componente gigante. La segunda implicación es que la remoción de los nodos esenciales debería ser más efectiva que la remoción de los nodos no esenciales. Como veremos a continuación, ninguna de las dos cosas sucede.

A. Desarme de la componente gigante

Una de las consecuencias de la hipótesis de Jeong es que las proteínas esenciales deberían tener un alto grado de centralidad, medida a través de distintos índices y, por ende, la remoción de los nodos con mayor centralidad debería ser más eficiente para el desarme de la componente gigante. En efecto, Zotenko y colaboradores mostraron que la betweenness es un mejor predictor de centralidad que el grado de los nodos. Por ende, estudiamos el desarme de la componente gigante de las cuatro redes, utilizando tres formas distintas de medir la centralidad de los nodos: grado, eigenvector y betweenness. Las primeras dos son medidas locales, asociadas con el comportamiento en la vecindad de cada nodo, mientras que la tercera es una medida global, relacionada con el rol de conectividad que los nodos desempeñan en la red. Ahora bien, si las proteínas esenciales que son los hubs realmente tienen un rol preponderante en la conectividad global de la red, esperaríamos que el criterio de sacar los nodos que tienen mayor betweenness fuera el más eficiente para el desarme de la componente gigante.

Para testear esto lo que hicimos fue tomar cada red y desarmar la componente gigante del siguiente modo: en cada paso de iteración se removió aquel nodo que tuviese el mayor valor de centralidad, según cada uno de los tres índices medidos. Luego el proceso se repitió sobre la componente gigante residual una y otra vez hasta desarmarla completamente. Además de usar los tres criterios de centralidad, se implementó un proceso random, el cual consistía en desarmar la componente gigante quitando nodos al azar. Los resultados del proceso se pueden consultar en la Fig. III A. Se puede advertir que todas las redes son robustas frente al desarme random. Esta propiedad ha sido atribuida en general a las distribuciones libres de escala [Jeong]. Sin embargo, se puede apreciar que el criterio más efectivo para el desarme no es la betweenness, sino el grado de los nodos. Esto no está de acuerdo con lo que esperábamos en caso de que la hipótesis de Jeong fuera correcta, ya que la betweenness es una medida global. Por el contrario, esta circunstancia sugiere que no lo es y que los hubs están asociados con efectos locales de su propio vecindario en lugar de estarlo con la conectividad global de la red. Este comportamiento inconsistente fue indicado no solamente por Zotenko y colaboradores, sino también por He y colaboradores.

B. Impacto diferenciado

Una segunda implicancia de la hipótesis de Jeong es que, si los nodos esenciales son efectivamente responsables de la conectividad de la red, esperaríamos que el desarme de la componente gigante fuera más eficiente al considerar solamente proteínas esenciales que al considerar un número equivalente de proteínas no esenciales. Es decir, esperaríamos que el impacto de remover hubs

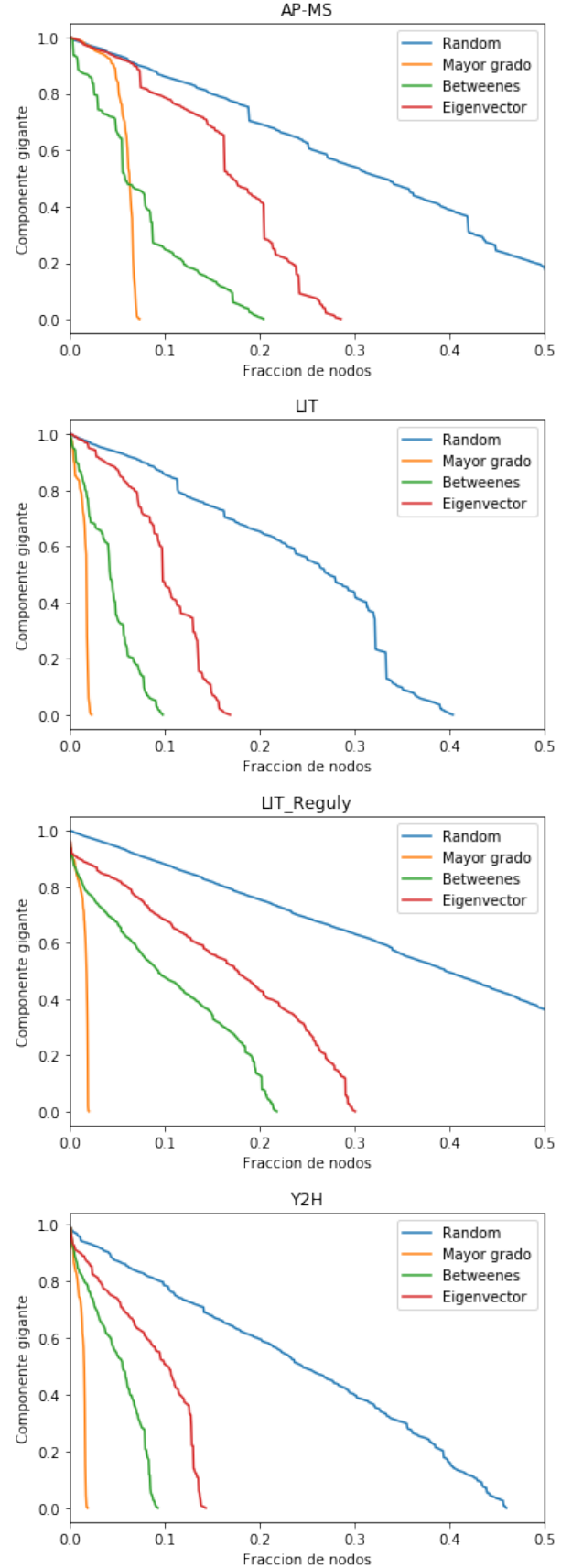


FIG. 2. Tamaño de la componente gigante en función de la proporción de nodos removidos de acuerdo con distintos criterios de centralidad de nodos en la red.

fuera más grande para nodos esenciales que para nodos no esenciales.

Para poner a prueba esta idea lo que hicimos fue una variante de la idea de Zotenko. Para cada una de las redes consideradas, desarmamos la componente gigante de dos maneras distintas. Por un lado, la desarmamos quitando un número dado de proteínas esenciales al azar. Por otro, la desarmamos quitando nodos no esenciales al azar, pero respetando la misma distribución de grado que tenían las proteínas esenciales removidas. Este procedimiento se repitió veinte veces para cada red. En la Fig. 3 se puede ver un gráfico del tamaño de la componente gigante residual en función del número de proteínas removidas al azar, fueran esenciales (rojo) o no esenciales (azul), para cada una de las cuatro redes, donde cada curva corresponde a una repetición. Se puede ver claramente que la remoción de los nodos esenciales es tan eficiente como la remoción de nodos no esenciales. Más aún, en la red LIT_Reguly los nodos no esenciales son más eficientes que los esenciales. Por consiguiente, esto suministra evidencia adicional contraria a la hipótesis de Jeong.

IV. ESENCIALIDAD: MÓDULOS BIOLÓGICOS VS. INTERACCIONES ESENCIALES

La hipótesis de He pretendía explicar la regla de Centralidad-letalidad sin apelar a la arquitectura de la red proteica. Para eso, postulaba la existencia de enlaces esenciales, distribuidos aleatoriamente a través de la red de interacciones. A partir de esta idea, He y sus colaboradores desarrollaron un modelo que predice la probabilidad de que una proteína al azar resulte esencial, en términos de:

1. La probabilidad de que los enlaces de la proteína sean esenciales (lo cual depende indirectamente del grado k de la proteína en cuestión)
2. La probabilidad de que la proteína resulte esencial por otros motivos como pueden ser enlaces esenciales con moléculas no proteicas.

La ecuación que obtienen para la probabilidad de que una proteína sea esencial es

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

He estimó α y β utilizando dos métodos alternativos, cuyos resultados demostraron ser consistentes incluso para distintas redes (es decir, los valores de cada red, determinados por distinto método coincidían razonablemente, aunque difirieran de una red a otra). En este trabajo, lo que hicimos fue utilizar uno de los métodos empleados por He: graficar la fracción de nodos esenciales en función del grado, linealizar la ec. 1, realizar un ajuste por mínimos cuadrados y calcular α y β a partir de

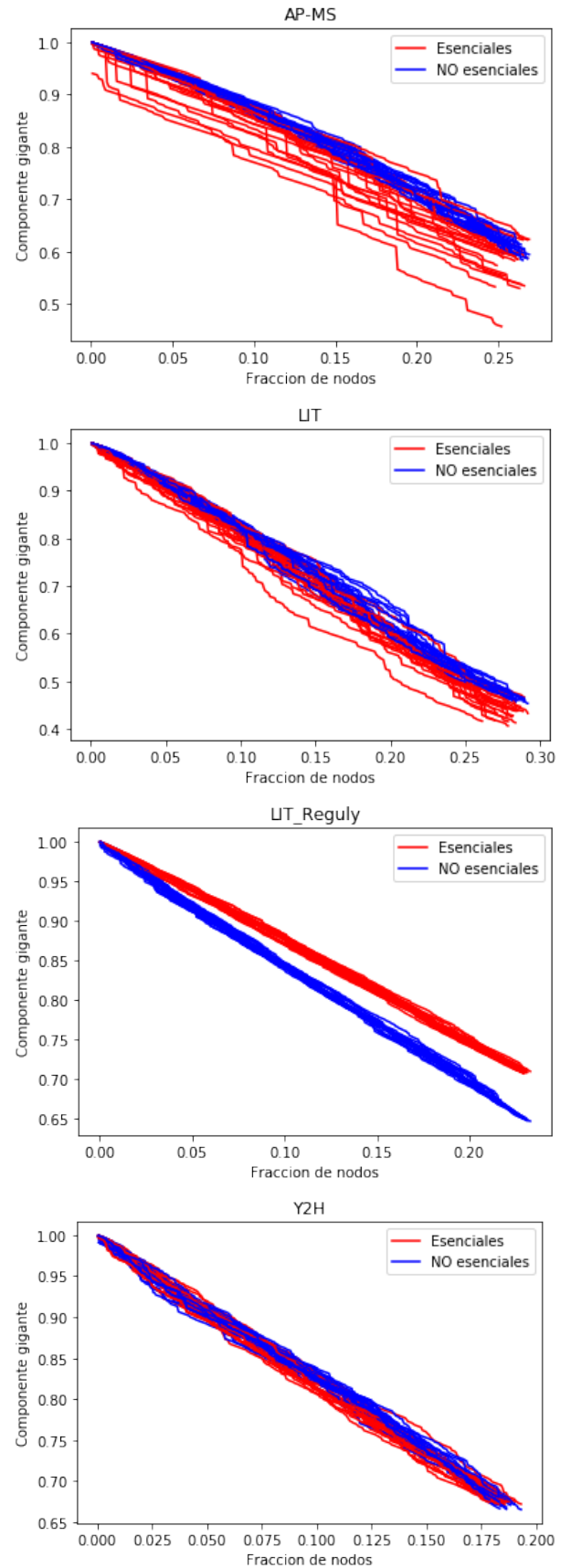


FIG. 3. Tamaño de la componente gigante residual en función del número de nodos removidos al azar, según sean esenciales (rojo) o no esenciales (azul), cada curva representa una repetición.

la pendiente y la ordenada al origen de la recta de ajuste. Los resultados del ajuste para cada una de las redes se muestran en la Fig. 4, mientras que los valores de α y β se resumen en la tabla III. Los valores difieren de una red a otra, aunque dos de estas redes tienen valores cercanos entre sí. En particular, los valores más pequeños están de acuerdo con los reportados por He y colaboradores, especialmente para la red LIT_Reguly.

	α	β
AP-MS	0.25	0.036
LIT	0.29	0.069
LIT_Reguly	0.079	0.045
Y2H	0.18	0.017

TABLA III. Valores de los coeficientes α y β obtenidos del ajuste por cuadrados mínimos, para cada una de las cuatro redes.

Sin embargo, Zotenko señaló que el modelo de He parte de una hipótesis central: se asume que la probabilidad de que una dada proteína sea esencial es independiente de la probabilidad de ser esencial de cualquier otra proteína no enlazada con ésta. Si fuera cierto, entonces esperaríamos que no haya correlaciones entre la cantidad de pares de proteínas esenciales y la cantidad de pares de proteínas no esenciales. Es decir, la cantidad de pares predichas por el modelo de He debería ser similar a la cantidad de pares efectivamente encontrados en la red.

Para testear esta idea tomamos las redes originales y calculamos la cantidad de pares de proteínas no enlazadas, tales que tuvieran al menos tres vecinos comunes (formaran parte de un mismo grupo) y que compartieran el mismo grado de esencialidad (ambas esenciales o ambas no esenciales). Luego, se utilizaron los α y β , estimados en el punto anterior, para determinar qué cantidad de proteínas con igual grado de esencialidad cabría esperar si el modelo de He fuese correcto. La cantidad de estos pares se determinó simplemente como la esperanza de que cada par fuera del mismo tipo, donde la probabilidad de que un dado par de proteínas 1 y 2 sean del mismo tipo estará dada por:

$$Prob_{mismo-tipo} = P_{E_1}P_{E_2} + (1 - P_{E_1})(1 - P_{E_2}) \quad (2)$$

Donde $P_{E_{1,2}}$ dependerán de α , β y del grado de cada proteína de acuerdo con la ec. 1.

Los resultados comparados se pueden encontrar en la tabla IV. Se puede observar que la cantidad de pares de proteínas con el mismo grado de esencialidad que efectivamente se encontró en la red es significativamente menor que la cantidad de pares predichos por el modelo de He para las redes AP-MS y LIT, mientras que la relación se invertía para las redes Y2H y LIT_Reguly. Estos resultados contradicen lo encontrado por Zotenko, si bien Zotenko no analizó exactamente el mismo tipo de redes y los resultados coinciden en la red Y2H, que sí tenemos en

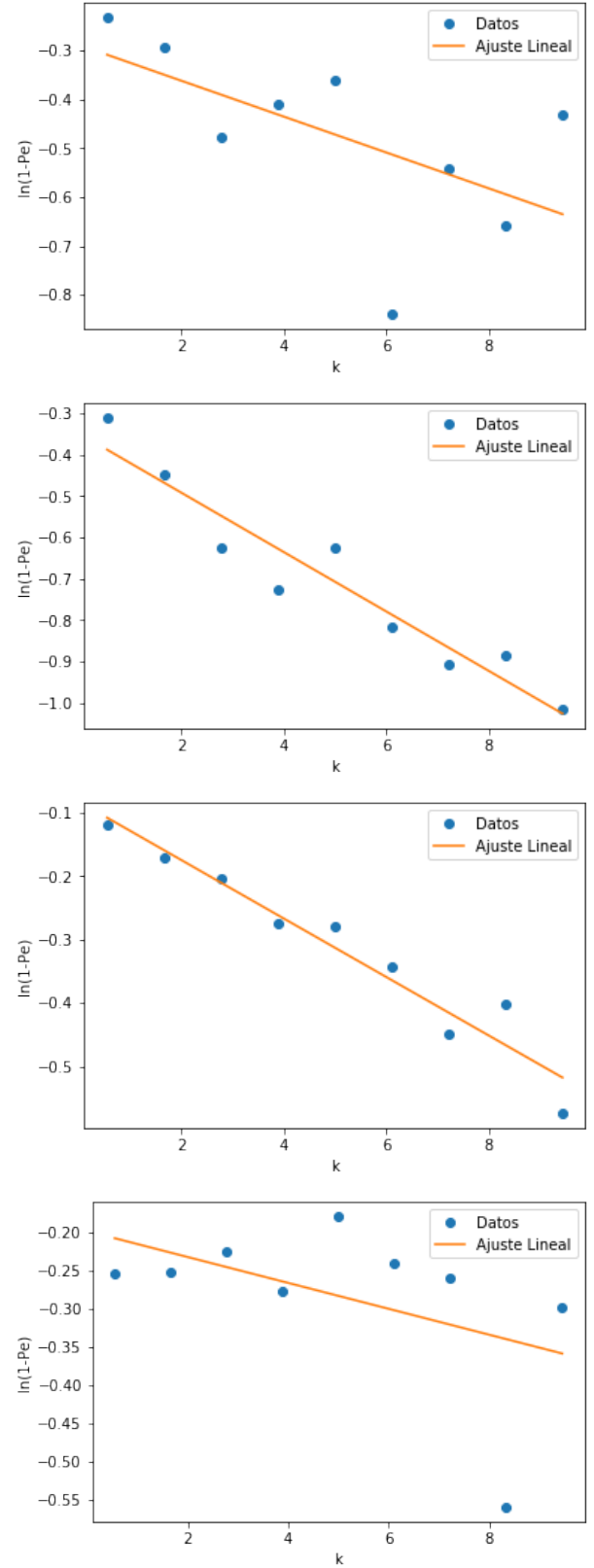


FIG. 4. Fracción de nodos no esenciales en función del grado de cada uno de los nodos, visualizado en escala logarítmica. La recta de color naranja corresponde al ajuste por mínimos cuadrados empleando la versión linealizada de la ec. 1.

común con dicho trabajo. Como consecuencia de lo antedicho, no es posible refutar el modelo de He, ni tampoco corroborarlo, ya que la evidencia no es concluyente.

	# total de pares	# de pares del mismo tipo	# Esperado (ajuste lineal)
AP-MS	12646	6957	11951
LIT	1392	1055	1099
LIT_Reguly	12251	7660	7414
Y2H	978	808	632

TABLA IV. Cantidad de pares con igual grado de esencialidad (ambas esenciales o ambas no esenciales) tanto para las redes reales como predichas por el modelo de He y colaboradores.

V. CONCLUSIONES

Se estudiaron cuatro redes de proteínas de la levadura *Saccharomyces cerevisiae*, a efectos de revisar

críticamente dos propuestas de explicación para la regla de centralidad y letalidad. En el caso del modelo propuesto por Jeong y colaboradores, se determinó que no hay una relación directa entre esencialidad y conectividad, puesto que medidas más sofisticadas de conectividad como la betweenness no son tan buenos predictores de esencialidad como el grado de los nodos. De hecho, los nodos esenciales y no esenciales no difieren en cuanto a su eficiencia en el desarme de la componente gigante. Esto sugiere que la esencialidad está asociada con un efecto local del nodo sobre sus vecinos.

En cuanto al modelo de He y colaboradores, se determinó que partía de la hipótesis de que la probabilidad de que una proteína fuera esencial era independiente de la probabilidad que tuvieran cualesquiera proteínas no enlazadas con ella. Al testear esta hipótesis se determinó que la cantidad de pares de proteínas no enlazadas con igual grado de esencialidad en la red real era significativamente inferior a la que predecía el modelo de He para la mitad de las redes, pero la relación se invertía para la otra mitad. En consecuencia, la hipótesis no pudo refutarse ni aceptarse por falta de evidencia concluyente al respecto.

-
- [1] Jeong H., Mason S. P., Barabási, A. L., Oltvai Z. N. *Lethality and centrality in protein networks*. Nature Brief Communications, 2001 May; 411:41-42.
 - [2] He X., Zhang J. *Why Do Hubs Tend To Be Essential in Protein Networks?* PLoS Genetics. 2006 Jun; 2(6):e88.
 - [3] Zotenko E., Mestre J., O’Leary D. P., Przytycka T. M.

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. 2008 Aug; 4(8):e1000140.