

Trabajo Práctico Computacional 02

Redes Complejas 2018

Noelia Parzajuk, Sofía Nicoletti

noelparzajuk@gmail.com - nicolettisofia1@gmail.com

17 de octubre 2018

1. Introducción

En el presente trabajo reproduciremos algunos de los resultados obtenidos en los trabajos de He et al. (2006) [2] y Zotenko et al. (2008) [1] para un conjunto de redes de levadura *Saccharomyces cerevisiae*. Dichos trabajos proponen modelos para explicar por qué la regla de centralidad-letalidad es aplicable en este tipo de redes. Esta regla se traduce como un fenómeno que relaciona la remoción de proteínas altamente conectadas y la supervivencia del organismo.

Caracterizaremos la estructura de las redes, la relación entre la conectividad del nodo y la esencialidad del mismo y el impacto de la remoción de nodos en relación diferentes criterios de centralidad. Finalmente, reproducimos uno de los tests que llevaron a cabo Zotenko et al. para rechazar el modelo que propone la investigación de He et al. En términos generales, analizaremos mediante distintos enfoques la relación entre la topología de las redes y las características fenotípicas del sistema biológico que representan.

2. Características de las redes

Características estructurales

Analizamos cuatro conjuntos de redes de interacción de la levadura *Saccharomyces cerevisiae*: Literature Curated (LIT y LIT Reguly), Yeast-two-hybrid (Y2H) y APMS (TAP-MS), a las cuales le agregamos el atributo de esencialidad, según corresponda, a aquellas asociadas a la supervivencia del organismo, y **removimos los nodos que poseían loops**. En la Tabla 1, presentamos las características generales de las redes con las que trabajaremos, indicando número de nodos y enlaces, grado medio y coeficiente de clústering medio.

Red	Nodos	Enlaces	Grado medio	Coeficiente de clústering medio
Y2H	1647	2518	3.05	0.05
LIT	1213	2556	4.21	0.32
LIT Reguly	3224	11291	7.01	0.26
APMS	1004	8319	16.57	0.64

Tabla 1: Propiedades estructurales de las redes de interacción de proteínas.

La Tabla 2 detalla la superposición de datos entre redes, es decir la fracción de nodos en común.

Explicar cómo se lee la tabla de superposición?

Entiendo que ren

Qué conclusiones pued

Ojo! la tabla de Zotenko es de la fracción de enlaces no de nodos!

Y2H	0.29	0.65	0.20
0.39	LIT	0.98	0.44
0.33	0.37	LIT Reguly	0.28
0.3	0.53	0.91	APMS

Tabla 2: Proporción de superposición entre redes.

Centrality-Lethality rule

supervivencia de la levadura dado que se remueve la p

En el trabajo de Jeong et al. [3], observó que el grado de un nodo en una red de interacciones de levadura está correlacionado con el efecto fenotípico de su remoción. Más aún, observó que nodos de alto grado (hubs) tienen mayor probabilidad de ser esenciales, es decir que están relacionados con la **supervivencia de la proteína**. Éste fenómeno es conocido como la regla *centralidad-letalidad*.

En consecuencia, resulta útil definir a partir de qué grado k es más probable que encontremos un nodo esencial. Para ello, nos preguntamos la fracción de hubs esenciales que podíamos encontrar si definíamos que a partir de cierto k , los nodos eran considerados como de alta conectividad. En la Figura 1 exponemos los resultados que obtuvimos. Se puede ver que en las redes, si el 20 % de los nodos son considerados hubs, el 60 % de ellos serán esenciales (Reproduciendo los resultados obtenidos por Zotenko et al.)

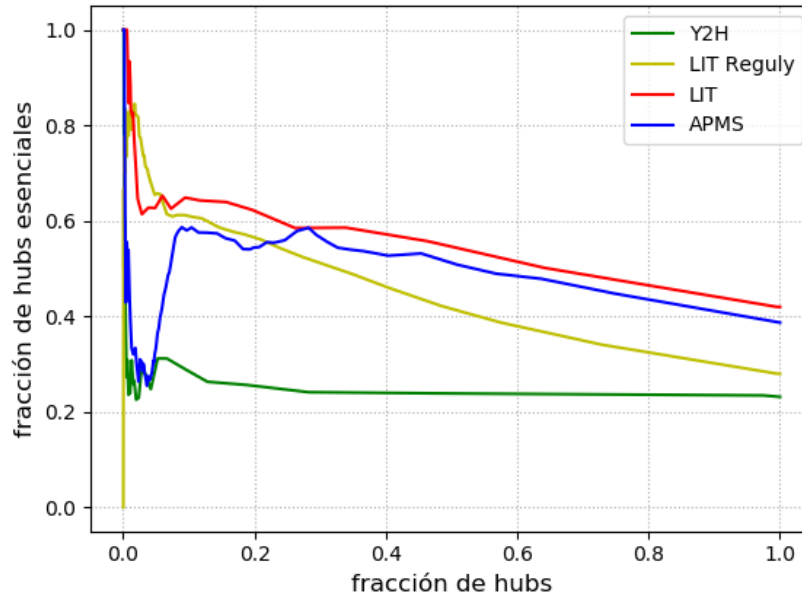


Figura 1: Fracción de hubs esenciales en función de todos los hubs, según el grado a partir del cual hayan sido definidos.

3. Análisis de la vulnerabilidad

Centralidad

El índice de centralidad de una red consiste en asignar un valor numérico a cada nodo de modo tal que cuantifique de alguna manera su relevancia en la topología de la red. Como verán, existen diversos índices de centralidad, donde cada uno adopta un criterio diferente y hace énfasis ciertos aspectos topológicos de la red.

Nosotras vamos a trabajar con cuatro centralidades: degree, shortest-path betweenness, eigenvector y subgraph. La primera es el caso más trivial: la centralidad de un nodo es igual a la cantidad de primeros vecinos que tenga. Por otro lado, las centralidades del tipo betweenness o de intermediación asignan a cada nodo un valor de centralidad basado en el rol que cumple éste para matener la conexión entre pares de nodos de la red. En particular, la centralidad shortest-path asigna a cada nodo un valor proporcional a la cantidad de caminos más cortos que pasan por él. Asimismo, la centralidad eigenvector se basa en qué tan cercano es un nodo a otros nodos centrales, es decir: una centralidad eigenvector alta, significa que un nodo está conectado con muchos otros nodos que también tienen a su vez una centralidad alta. Por último, la centralidad subgraph asigna a cada nodo un valor de centralidad basándose en el número de caminos cerrados que se pasan por él.

Conectividad y desarme

Conociendo los diferentes criterios de centralidad resulta interesante preguntarse qué rol cumplen los nodos centrales en la conectividad de la red. Naturalmente pensamos en los *hubs*, los nodos con más alta centralidad de grado. La remoción de alguno de ellos puede afectar fuertemente la capacidad de otros nodos para conectarse entre sí, pero ¿qué tanto afecta? Probablemente más que quitar un nodo al azar. Este tipo de análisis es el que abordaremos en esta sección, nos preguntamos qué tan efectivos resultan los nodos de alta centralidad (bajo diferentes criterios) al momento de “desarmar” una red, qué relación se establece ente centralidad y conectividad. Para ello, iremos removiendo progresivamente los nodos más centrales de cada red y viendo su impacto en la conectividad de la red.

Una forma posible de medir el impacto de la remoción de nodos en la conectividad de una red es calculando el decrecimiento de la componente más grande. En la Figura 2 se muestra cómo afecta la remoción de nodos centrales, random y esenciales en la conectividad de la red, para las cuatro redes de interacción de proteínas. La curva random se realizó promediando diez veces.

Vemos en la Figura 2 que la eliminación de nodos más centrales usando el criterio de intermediación (shortest-path betweenness) es el que tiene mayor impacto en las cuatro redes. El desarme de la red allí se produce de manera más abrupta que con los demás criterios de centralidad. De todas maneras, notamos que remover nodos con algún criterio de centralidad es mucho más disruptivo que la eliminación de proteínas de manera aleatoria. La degree resulta levemente más eficiente en el desarme que eigenvector y subgraph, aunque cualitativamente vemos que estas tres siguen una misma tendencia: menos disruptivas que betweenness pero mucho más que random. Esto no resulta casual si consideramos que degree, eigenvector y subgraph son centralidades denominadas *locales*, donde su valor depende principalmente de sus inmediaciones, mientras que shortest-path mide la importancia de un nodo en la conexión pares de nodos de toda la red.

En cuanto a la vulnerabilidad de las redes notamos que APMS se presenta como la más robusta frente a la remoción de proteínas centrales, mientras que Y2H es indiscutiblemente la más débil. Esto es consistente con el método de relevamiento y asignación de enlaces entre proteínas para estas redes.

Por que es consistente? Que aspecto

Grado y esencialidad en la conectividad

Posteriormente analizamos el rol de los hubs en la conectividad y nos preguntamos si su eficacia en el desarme de la red es mayor si son esenciales. En primer lugar, notemos que la remoción de todas las proteínas esenciales es menos disruptivo para la conectividad que remover la misma fracción de nodos de acuerdo a algún índice de centralidad, incluyendo la de grado (ver Figura 2). Este patrón está presente en las cuatro redes.

En segundo lugar, como se muestra en la Tabla 3, se comparó el impacto de la eliminación de nodos esenciales contra la eliminación de nodos aleatorios no-esenciales preservando la distribución de grado. Nuevamente, el impacto se mide en relación a la cantidad de nodos que conserva la componente más grande respecto de lo que tenía al comienzo. Debido a que, por lo general, no es posible encontrar nodos no-esenciales con exactamente la misma distribución de grado que los esenciales (pues muchos de ellos son hubs), se removieron nodos con la distrubución más parecida posible. Esto es: se escoge de manera aleatoria un nodo no-esencial cuyo grado sea lo más cercano posible al grado del esencial. El valor presentado en la tabla es el promedio de 100 iteraciones, junto con su desviación estándar (notemos que es pequeña, pues no hay muchas maneras de escoger nodos no-esenciales que

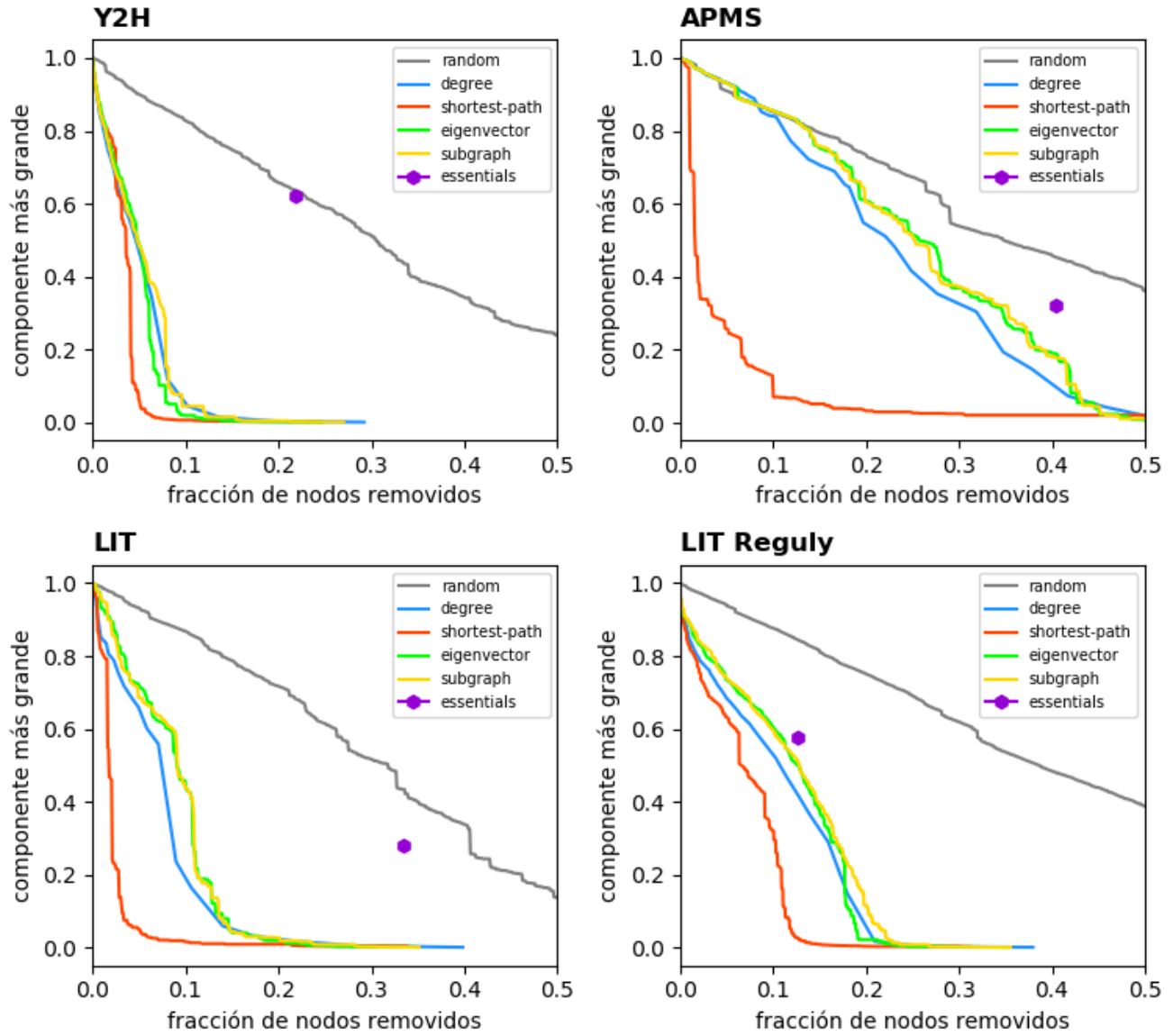


Figura 2: **Vulnerabilidad frente a la remoción de proteínas centrales.** El impacto de la remoción de nodos de la red es cuantificado por la fracción de nodos de la componente más grande. Hay una curva para cada criterio de centralidad, que muestra la fracción de nodos de la componente más grande en función de la fracción de nodos más centrales removidos. También se muestra el impacto de remover nodos de manera aleatoria y el tamaño de la componente más grande si se remueven todas las proteínas esenciales.

preservan la distribución de grado). Vemos que, a excepción de la red LIT, la remoción de proteínas esenciales no es más disruptiva que la remoción de nodos no-esenciales equivalentes en cuanto al grado.

Concluimos que, si bien la remoción de nodos de alto grado afecta fuertemente a la conectividad de la red, esto no se relaciona con el carácter de esencialidad de los mismos.

Red	Esenciales	Random no-esenciales
Y2H	0.624	0.625 ± 0.011
LIT	0.281	0.414 ± 0.004
LIT Reguly	0.575	0.540 ± 0.004
APMS	0.324	0.374 ± 0.014

Tabla 3: El impacto de la eliminación de nodos esenciales comparado con la eliminación de nodos no-esenciales con la misma distribución de grado. El impacto en la conectividad de la remoción de proteínas se mide en términos de número de nodos de la componente más grande normalizado por el número de nodos de la componente gigante original.

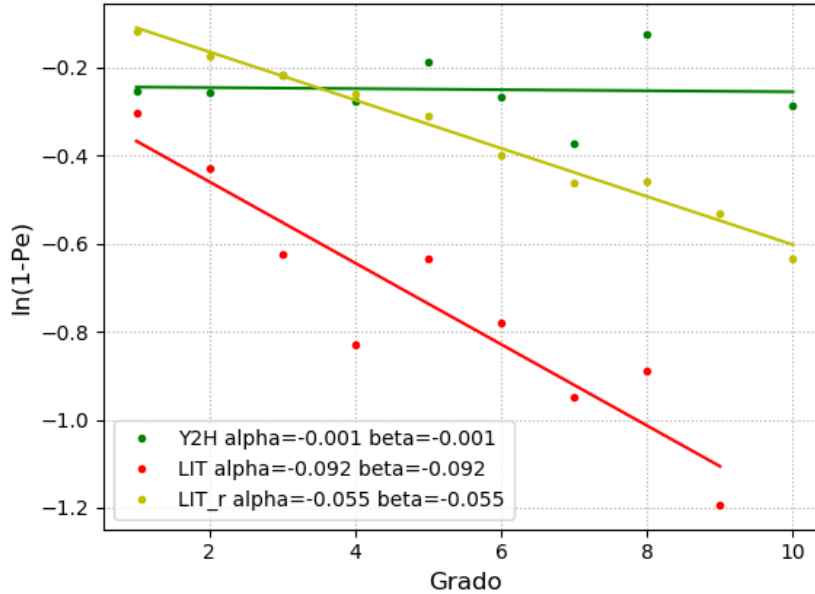
4. Esencialidad: Módulos biológicos vs Interacciones Esenciales

Por la dificultad que implica identificar interacciones esenciales entre proteínas (PPIs) experimentalmente, He et al. propone modelar dicha cantidad teniendo en cuenta que dos proteínas que forman un enlace esencial deben ser esenciales, mientras que un enlace esencial entre proteínas no implica que estas lo sean.

En particular, estudiaremos la probabilidad de que una proteína sea esencial, en función del grado de la misma. A partir de dicho valor, podremos calcular la probabilidad α de que un enlace entre proteínas sea esencial. Para que una proteína no sea esencial, por un lado, no tiene que tener PPI esenciales y, por el otro, no debe ser afectada por factores externos que le aporten la esencialidad. Es así como la probabilidad (P_e) de que una proteína con grado k sea esencial es:

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

donde β aporta la posibilidad de que una proteína sea esencial por otros factores independientes de pertenecer a un enlace con dicha característica. De acá se sigue que $\log(1 - P_e)$ varía linealmente con k . La figura 3 detalla la evolución de dicha probabilidad que observamos en las redes, con el correspondiente ajuste lineal que nos permitió calcular las probabilidades α y β .



Ojo que aca la ordenada al or

Figura 3: Ajuste lineal entre $\ln(1 - P_e)$ y el grado del nodo.

Zotenko et al. rechaza el modelo de interacciones de He et al., a partir de las hipótesis que él realiza. Notó que, si dos proteínas no interactúan, luego la esencialidad de una proteína en tal par no depende de la esencialidad de la otra proteína. Además, esta independencia debería ser también observada incluso cuando las proteínas comparten vecinos.

Para comprobar el resultado de Zotenko, contamos los pares de proteínas no adyacentes que poseen 3 o más vecinos en común y cuántos de estos pares son del mismo tipo, es decir ambas esenciales o no esenciales. Luego, comparamos éste último resultado con el valor esperado del mismo a partir del modelo de He. Éste último se calcula a partir del siguiente modelo:

$$\langle \text{Pares iguales} \rangle = \left(\sum_{\text{pares}} P_{k_1} P_{k_2} + (1 - P_{k_1})(1 - P_{k_2}) \right) / N \quad (2)$$

donde P_{k_i} es la probabilidad de que un nodo de grado k_i sea esencial y N es el número total de pares. La red APMS fue excluida del análisis dado que el modelo de He et al. puede no funcionar para redes en las cuales las proteínas pertenezcan a "subconjuntos".

El total de pares del mismo tipo esperado por el modelo es significativamente diferente del observado. En consecuencia, la interacción entre dos proteínas no está correlacionada con que ambas compartan una característica (esencialidad o no)

Red	Pares	Pares del mismo tipo	Pares del mismo tipo esperados
Y2H	522	352	86
LIT	718	383	91
LIT Reguly	10777	6187	1811

Tabla 4: Diferencia entre los pares del mismo tipo esperados y los observados.

Las conclusiones tienen sentido, aunque me extraña que los números sean tan b

Referencias

- [1] *Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality* (2008) Zotenko E., Mestre J., O'Leary D. P., Przytycka T. M, PLoS Computational Biology, Volume 4, Issue 4.
- [2] *Why Do Hubs Tend to Be Essential in Protein Networks?* (2007) He X., Zhanh J., PLoS Computational Biology, Volume 2, Issue 6.
- [3] *Lethality and centrality in proteing networks* (2007) Jeong H., Mason S.P., Barabási A-L, Oltvai Z. N. Nature, Vol 411.