

# Centralidad-Letalidad: un camino hacia la mesoescala

Ferreira Chase, T., Leizerovich, M., Goren, G.

17 de octubre de 2018

## Resumen

En este trabajo se imitan algunos de los resultados obtenidos por He et al.(2006) y Zotenko et al.(2008) referidos a la llamada *regla de centralidad-letalidad*, según la cual los nodos de mayor grado (*hubs*) en redes de interacción biomolecular tienden a corresponder a proteínas biológicamente esenciales (i.e. necesarias para la supervivencia del organismo). En particular, se estudiaron cuatro redes de interacciones de proteínas correspondientes a la levadura *Saccharomyces cerevisiae*. Se realizó un análisis de la hipótesis según la cual la esencialidad de los *hubs* se debe a su rol en la conectividad de la red, siendo esta rechazada, y se procedió a considerar el modelo de interacciones esenciales propuesto por He. Este modelo resultó relativamente consistente pero, tal como se observó en dos de las cuatro redes, no es capaz de capturar las correlaciones presentes entre proteínas esenciales no adyacentes.

## 1 Introducción

En el estudio de redes biomoleculares (de genes, de proteínas, etc.), uno de los principales desafíos consiste en caracterizar el vínculo entre, por un lado, la funcionalidad biológica de las moléculas involucradas, y por el otro la estructura topológica de la red que conforman, i.e. la información sobre el sistema biológico accesible a través del estudio del cableado o patrón de interacciones entre dichas moléculas. En este contexto, es relevante preguntarse por la posible correlación entre nociones biológicas y topológicas de *importancia*. La llamada *regla de Centralidad-Letalidad* afirma que los

nodos de grado más alto en una red (i.e. los *hubs*) tienden a ser más esenciales biológicamente que lo esperado por azar. Aquí, por “esencial” entendemos que dicha proteína, en experimentos en los que el organismo es libre de sobrevivir y reproducirse sin presiones externas, es indispensable para la supervivencia del organismo (una remoción de la misma impide la supervivencia). Esta regla constituye una observación experimental en sí misma.

La regla de Centralidad-Letalidad fue planteada por primera vez por Jeong et al. en el 2001 [1]. Una primer análisis del fenómeno puede sugerir fuertemente una conexión con el hecho de que los *hubs* son más importantes que los *no-hubs* en cuanto al mantenimiento de la estructura global de la red [2]. Es decir, la remoción de un *hub* genera un impacto mayor en la estructura de la red (ya sea considerando la conectividad de la componente gigante o el diámetro de la red) que la remoción de un nodo aleatorio. Esto muestra que los *hubs* ocupan un lugar de gran importancia desde un punto de vista puramente topológico, y resulta natural entonces hipotetizar que esta correlación entre dos nociones de importancia (biológica y topológica) es en realidad una relación de causalidad. Bajo esta hipótesis, la supervivencia del organismo depende de la conectividad de las distintas regiones de la red. Esta fue la explicación planteada por Jeong et al., y analizaremos la validez de la misma en la sección 3.

Por otro lado, He propuso en un trabajo[2] que la esencialidad de los nodos no está dada por una estructura a gran escala de la red, sino que está distribuida de cierta manera aleatoria, en la cual la preponderancia de grados altos entre proteínas esenciales es solo una conse-

cuencia de un mecanismo más neutro. La parte esencial de este modelo, denominado "modelo de interacciones esenciales", es que en él los nodos son esenciales dado que forman parte de una o más interacciones proteína-proteína esenciales, o por algún factor externo a la red de interacción de proteínas (por ejemplo, porque forman parte de una interacción proteína-gen esencial). Luego, los *hubs* tienden a ser esenciales ya que al tener una gran cantidad de vecinos, tienen más probabilidades de participar de una interacción esencial. A la vez, el autor propone un modelo para calcular la probabilidad de que un nodo sea esencial, el cual sólo depende del grado del nodo y no existen correlaciones entre dos nodos no adyacentes. Es decir, la probabilidad de que un nodo sea esencial sí depende de si sus vecinos lo son, dado que existe la posibilidad de que lo sean debido a enlaces esenciales que incluyan a la proteína en cuestión. Pero, siempre bajo este modelo, sí son independientes las probabilidades de ser independientes de dos proteínas no adyacentes.

	# Nodos	# Enlaces	Grado medio	Clustering medio
<b>Lit</b>	1213	2621	4.32	0.33
<b>Lit_reg</b>	3224	11789	7.31	0.27
<b>APMS</b>	1004	8319	16.57	0.65
<b>Y2H</b>	1647	2682	3.26	0.06

Cuadro 1 – Tabla con las características fundamentales de las redes LIT, LIT-REG, Y2H y APMS.

Por último, Zotenko mostró en un trabajo[3] que los *hubs* no solo no son los nodos más importantes a la hora de mantener la conectividad de la red, sino que la probabilidad de que un nodo sea esencial no es, en general, independiente de la probabilidad de otros nodos no vecinos lo sean. En particular, Zotenko ofrece una nueva explicación para la esencialidad de los nodos. En su trabajo muestra que dado que los nodos esenciales no están distribuidos de forma aleatoria en las distintas subredes *densas*, y dado que las subredes se forman a partir de nodos que comparten características biológicas, entonces los módulos que contienen mayoría de nodos esenciales son los que llevan el carácter

de esencialidad. Éstos módulos se denominan *Módulos complejos biológicos esenciales* (ECOBIMs).

## 2 Caracterización de las redes de interacciones analizadas

Para nuestro estudio, vamos a analizar cuatro redes de interacciones de proteínas de la levadura *Saccharomyces cerevisiae*. Dos de ellas fueron compiladas a partir de datos en literatura, una de ellas desarrollada por Regul y et al. Estas dos redes son denominadas Lit y Lit-Reg. En algunos trabajos se sugirió que, debido a su importancia biológica, las proteínas esenciales suelen tener mayor relevancia en las redes de literatura y por lo tanto, a la hora de crear una red como las analizadas, estas proteínas podrían tener un grado mayor de lo esperado. Por este motivo se analizaron dos redes con un alto componente de datos experimentales: una red Yeast Two Hibrid (Y2H), que contiene una gran componente de interacciones que fueron detectadas experimentalmente, y una red que contiene interacciones por encima de un umbral de confianza prefijado (APMS). Se sabe que esta red sobreestima la cantidad de enlaces reales debido a que representa como una *clique* a cada conjunto de proteínas que se observó interactúan con una dada proteína “anzuelo”.

A partir de aquí, Zotenko et al. analiza únicamente la componente gigante de cada red de interacciones. En este trabajo siempre se analizará la componente gigante de la red a menos que se especifique lo contrario.

En el Cuadro 1, se sintetizan las características más importantes de dichas redes, mientras que en el Cuadro 2 se presentan los solapamientos entre los conjuntos de proteínas considerados en cada red.

Para exponer la correlación positiva entre la esencialidad de los nodos de la red y su grado, podemos preguntarnos cuándo la fracción de nodos esenciales en *hubs* es mayor que la que cabría esperar en el promedio de la red. Para ello, se introdujo un criterio móvil (umbral) para definir cuándo consideramos que un nodo es

o no es un *hub*. Haciendo un barrido sobre dicho umbral uno puede graficar la fracción de nodos *hubs* esenciales en función de los nodos *hubs* totales para diferentes valores del umbral. El resultado puede observarse en la Figura 1.

Tal como puede observarse, se verifica que a medida que se incluyen en la cuenta nodos de grado más bajo, disminuye la fracción total de nodos esenciales<sup>1</sup>. Esto constituye una verificación de la regla de Centralidad-Letalidad.

	Lit	Lit_reg	APMS	Y2H
Lit	1.00	0.98	0.37	0.07
Lit_reg	0.22	1.00	0.17	0.03
APMS	0.12	0.25	1.00	0.01
Y2H	0.07	0.15	0.04	1.00

Cuadro 2 – En esta tabla se presentan las fracciones de enlaces de las redes de la columna izquierda, contenidas en las redes de la primera fila. Por ejemplo, el 98 % de los enlaces presentes en LIT se encuentran también en LIT REG.

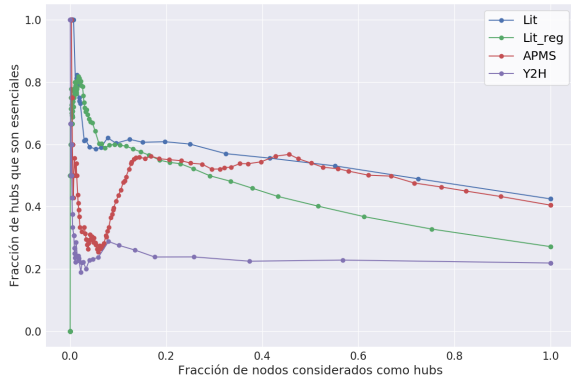


Figura 1 – Proporción de nodos esenciales hubs versus proporción de *hubs* en la componente gigante de la red.

<sup>1</sup>Una excepción a este fenómeno parece ocurrir para la red APMS cuando se consideran umbrales altos (fracción de nodos considerados como esenciales pequeña). Sin embargo esto puede deberse a fluctuaciones asociadas con considerar una cantidad baja de nodos dentro de la cuenta, y la tendencia general se recupera para umbrales más bajos.

### 3 Análisis de vulnerabilidad

En esta sección se estudió cómo diferentes criterios de centralidad en las redes de proteínas capturan los puntos débiles de la misma, esto es su estructura de vulnerabilidad. Esto es, definiendo cuales son los nodos más importantes de la red y realizando un desarme secuencial, estudiar la velocidad de desarme de la componente gigante de la red.

Se analizaron las cuatro redes de proteínas utilizando seis criterios diferentes de centralidad para sus proteínas (degree, eigenvalues, subgraph, betweenness, current flow). En cada caso, se remueve un nodo por vez, eligiendo siempre el más central para la red actual, hasta desconectar por completo la red. Adicionalmente, se realizó este mismo procedimiento 100 veces eligiendo cada nodo de manera aleatoria, promediando los resultados de las 100 realizaciones.

En esta sección, se analizan las redes completas, en vez de analizar únicamente las componentes gigantes.

El criterio *degree* asigna mayor centralidad a los nodos de mayor grado. El criterio de centralidad por autovalor (*eigenvalues*) asigna mayor centralidad a nodos cuyos vecinos son más centrales, de una manera recursiva. Por otro lado, el criterio *betweenness* ordena los nodos según la fracción de caminos más cortos entre nodos que pasan por el nodo en cuestión.

La centralidad *subgraph* para el nodo  $n$  es la suma de los pesos de los caminos cerrados que empiezan y terminan en el nodo  $n$ . El peso del camino aumenta con el tamaño del mismo. Cada camino está asociado con un subgrafo conexo. Finalmente, la *current flow betweenness* utiliza un modelo de corriente eléctrica para calcular la centralidad (un nodo es más central si fluye más corriente por él).

Finalmente, se compararon los resultados con la fracción de la componente gigante que sobrevive ante la eliminación de las proteínas esenciales de la red (no de forma gradual sino en un único paso).

Como se puede observar en la Figura 2, en ningún

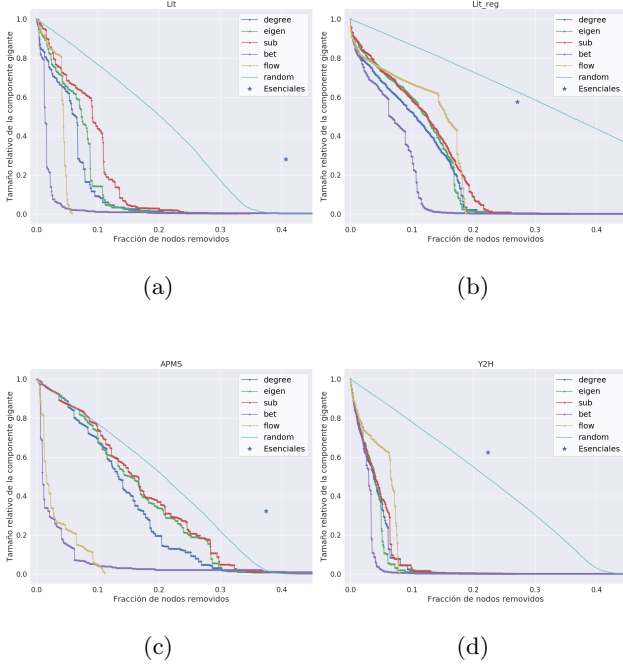


Figura 2 – *Fracción de la componente gigante sobreviviente en función de la fracción de nodos removidos, según los distintos criterios, para las 4 redes.*

caso la centralidad *degree* es la más adecuada para desarmar la componente gigante de la red. Esto podría sugerir que el grado no es una medida infalible de la importancia topológica de la red. En contraposición, la centralidad *betweenness* es la que resulta más efectiva para el desarme en todas las redes. Dada la importancia que asignan Jeong et al. al grado como medida de la importancia topológica, esto en sí mismo podría arrojar dudas sobre su hipótesis. Por otro lado se observa que la eliminación de nodos en orden aleatorio es muchísimo menos efectiva que cualquiera de los desarmses por centralidad.

Finalmente, se puede apreciar que la eliminación de los nodos esenciales disminuye la componente gigante en una fracción cercana a la que cabe esperar para los desarmses aleatorios. Sin embargo, dado que los desarmses para los criterios de centralidad se realizaron de manera secuencial en comparación con la eliminación de nodos esenciales que fue realizada simultáneamente, se propone el siguiente análisis para comprobar que no existe correlación entre el desarme aleatorio y la eliminación de nodos esenciales.

Para estimar la correlación entre la esencialidad de los nodos de la red y su centralidad, se procedió a comparar la fracción de la componente gigante que sobrevive a la eliminación de los nodos esenciales, y la fracción que sobrevive a eliminar de manera aleatoria la misma cantidad de nodos no esenciales que siguen la misma distribución de grado. Para este análisis se utilizaron dos criterios para la elección de nodos aleatorios no esenciales, en el caso que se debieran remover más nodos de este tipo de los que era posible. Sean  $n$  nodos esenciales de grado  $k$ . A la hora de remover los nodos no esenciales con la misma distribución de grado, cabe preguntarse qué ocurre cuando existen  $m$  nodos no esenciales de grado  $k$ , con  $m < n$ .

Primer criterio: Se busca el grado más cercano a  $k$  ( $\bar{k}$ ). Si la cantidad  $n(\bar{k}) = p$  es tal que  $m + p \geq n$ , entonces se eliminan  $n$  nodos de la “bolsa” formada por los grados de grados  $k$  y  $\bar{k}$ . En caso contrario, se siguen agregando nodos a la bolsa recorriendo grados cercanos al original y agregándolos a la “bolsa” hasta que el número de nodos sea mayor a  $n$ , y recién entonces se eligen  $n$  nodos aleatoriamente.

Segundo criterio: Primero, se eliminan los  $m$  nodos de grado  $k$ , por lo que aún deben eliminarse  $n - m$  nodos. A continuación, se procede a eliminar nodos de grado  $\bar{k}$ . Si  $p \geq n - m$ , se eliminan los nodos necesarios de grado  $\bar{k}$ . Si  $p < n - m$ , se eliminan todos los nodos de grado  $\bar{k}$ , se recalcula el grado a menor distancia de  $k$  y se repite el proceso para eliminar  $n - m - p$  nodos. Este proceso es iterativo hasta eliminar los  $n$  nodos necesarios.

El criterio 1 es intrínsecamente más aleatorio, mientras que el criterio 2 garantiza que todos los nodos eliminados tienen grado lo más cercano posible al grado del nodo esencial correspondiente.

Se observó que el segundo criterio resulta más adecuado en valor medio para que la eliminación de nodos no esenciales aleatorios se asemeje al resultado de la eliminación de nodos esenciales, sobre todo en la red APMS. Se produce un solapamiento de la red Y2H entre los dos valores. También se observa un solapamiento para el primer criterio para la red Lit. Se concluye que

	Esenciales	No esenciales equivalentes	Z-scores
<b>Lit</b>	0.28	0.437 +/- 0.008	-20.06
<b>Lit_reg</b>	0.58	0.565 +/- 0.007	1.47
<b>APMS</b>	0.32	0.434 +/- 0.021	-5.15
<b>Y2H</b>	0.62	0.621 +/- 0.011	0.26

	Esenciales	No esenciales equivalentes	Z-scores
<b>Lit</b>	0.28	0.415 +/- 0.004	-37.67
<b>Lit_reg</b>	0.58	0.539 +/- 0.004	8.44
<b>APMS</b>	0.32	0.396 +/- 0.013	-5.53
<b>Y2H</b>	0.62	0.622 +/- 0.011	0.16

Cuadro 3 – Tablas con la fracción de la componente gigante que sobrevive al eliminar nodos con la misma distribución de grados. En el primer caso se eliminan los nodos esenciales de la red y en el otro se eliminan nodos no esenciales elegidos de manera aleatoria. Para eliminar nodos aleatorios, se utilizaron dos criterios diferentes, uno que es intrínsecamente más aleatorio (arriba) y otro que es menos aleatorio y garantiza que todos los nodos eliminados tengan grado lo más cercano posible al grado del nodo esencial correspondiente.

para la red Y2H y la red Lit, podemos descartar que la esencialidad de las proteínas este correlacionada con su "letalidad"(topológica), aunque no podemos afirmar lo mismo para las redes APMS y Lit Reg.

## 4 De las proteínas esenciales a las interacciones esenciales

Según He, la esencialidad de los nodos se debe a que los mismos participan de enlaces esenciales entre proteínas. Luego, un nodo con alto grado tiene más probabilidades de participar de un enlace esencial, por lo que un *hub* tiene mayor probabilidad de ser esencial que un nodo de menor grado. Por otro lado, no todos los enlaces entre proteínas esenciales son esenciales, ya que la esencialidad del nodo puede estar dada por factores externos a la red (como interacciones proteínas-genes, etc.).

Sea  $\alpha$  la fracción de enlaces esenciales en la red, y sea  $\beta$  la fracción de nodos de la red que son esenciales

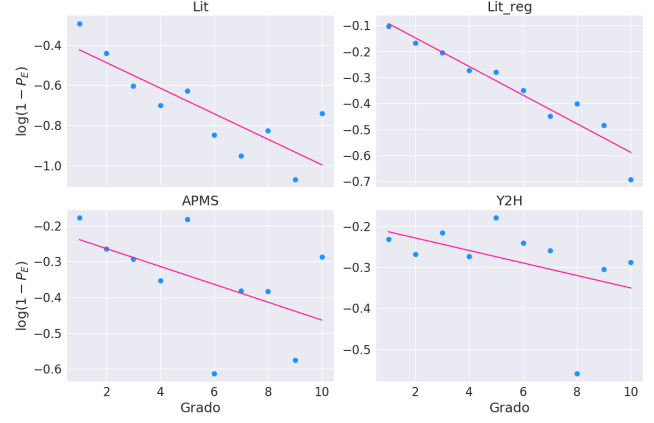


Figura 3 – En esta figura se puede observar la relación dada por la ecuación 2 entre la probabilidad de que un nodo sea esencial y el grado del nodo, para cada red.

debido a otros factores distintos de participar de enlaces esenciales. Luego, la probabilidad de que un nodo aleatorio de grado  $k$  sea esencial ( $P_E$ ) está dada por la ecuación 1

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

es decir, que el nodo participe de un enlace esencial con alguno de sus  $k$  vecinos, o que sea esencial debido a factores externos.

Luego, reescribiendo la ecuación 1 y tomando logaritmo natural a ambos lados, se obtiene una relación lineal (ecuación 2) entre el logaritmo de la probabilidad y el grado del nodo.

$$\ln(1 - P_E) = k \cdot \ln(1 - \alpha) + \ln(1 - \beta) \quad (2)$$

Para estudiar la dependencia del grado de los nodos con la probabilidad de que sea esencial, se consideró que  $P_E$  está dado experimentalmente por la fracción de nodos esenciales de grado  $k$ , con respecto a la cantidad total de nodos de grado  $k$ . Luego, si se grafica dicha dependencia (ver figura 3), se puede observar que todas las redes presentan una relación lineal que se puede ajustar mediante la ecuación 2.

Se estudió dicho comportamiento para grados menores a 10 ya que se observaron pocos nodos para mayores valores de  $k$ .

Los valores  $\alpha$  y  $\beta$  para las distintas redes se pueden obtener de los parámetros del ajuste lineal y comparándolos con los mismos en la ecuación 2. Dichos valores se pueden observar en la tabla 4.

	Pendiente	Ord. al or.	$R^2$ adj.	$\alpha$	$\beta$
<b>Lit</b>	-0.064 +/- 0.015	-0.361 +/- 0.094	0.651	0.062 +/- 0.014	0.303 +/- 0.065
<b>Lit_reg</b>	-0.055 +/- 0.006	-0.038 +/- 0.036	0.908	0.054 +/- 0.006	0.037 +/- 0.035
<b>APMS</b>	-0.025 +/- 0.015	-0.214 +/- 0.091	0.172	0.025 +/- 0.014	0.193 +/- 0.074
<b>Y2H</b>	-0.015 +/- 0.011	-0.198 +/- 0.068	0.096	0.015 +/- 0.011	0.180 +/- 0.056

Cuadro 4 – Se presenta la probabilidad de que un enlace sea esencial ( $\alpha$ ) y la probabilidad de que un nodo sea esencial debido a factores externos, para las 4 redes.

#### 4.1 De las proteínas esenciales a las interacciones esenciales - Analisis alternativo

Se propuso estudiar el número de enlaces esenciales, y su relación con la red mediante el primer método propuesto por He para las redes APMS, Y2H y LIT.<sup>2</sup> Este método consiste en *recablear* la red inicial manteniendo el grado de los nodos. Es decir, enlazar de forma aleatoria a la red, siempre manteniendo el número de vecinos por nodo.

Se considera un IBEP (interaction between essential proteins) a un enlace entre dos nodos esenciales. Luego, contando el número inicial de IBEPs y recableando la red 10000 veces, se generó un histograma de la cantidad de IBEPs por recableo (ver figura 4 para este análisis con la red APMS). De esta forma es posible observar la cantidad de IBEPs prevalecen en la red. Se observo que para ninguna red, el número de IBEPs de las simulaciones fue mayor al de la red original.

Si se supone que el excedente entre IBEPs en las simulaciones corresponden a proteínas esenciales, se puede calcular el porcentaje de proteínas esenciales ya

Distribucion de enlaces esenciales para recableos de la origini:

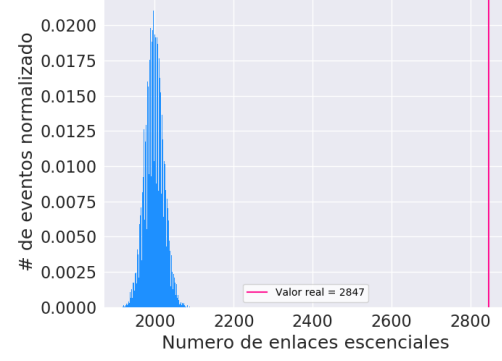


Figura 4 – Histograma generado a partir de 10000 recableos de la red APMS, contando el número de IBEPs para cada recableo. En rojo, el número de IBEPs en la red original.

que participan de enlaces esenciales como

$$\alpha = \frac{x - \langle x \rangle}{N}, \quad (3)$$

donde N es el número total de nodos.

Por otro lado, existen otros factores que pueden promover el carácter de esencial a una proteína, aparte de si su interacción con otra proteína es esencial. Para calcular la probabilidad de que un nodo sea esencial por otro factor aparte de PPI ( $\beta$ ), se realizaron simulaciones de recableo para las redes, quitándoles la información de si los nodos son esenciales por PPI o no. Para esto, se recableo la red inicial de la siguiente forma: se arrojaron  $(x - \langle x \rangle)$  enlaces esenciales de forma aleatoria en la red, y luego se fueron arrojando nodos marcados como esenciales de forma aleatoria (también teniendo en cuenta los que previamente fueron dotados de carácter esencial por tener un IBEP) hasta llegar al número de nodos esenciales original. Notar que  $(x - \langle x \rangle)$  es el número de enlaces esenciales que uno esperaría en la red original. Luego, se puede asociar esta fracción de nodos con respecto a N como la probabilidad  $\beta$ . Es decir, ésta fracción de nodos corresponde a la probabilidad de que un nodo sea esencial debido a factores externos de la red, y no a que participa de un enlace esencial. En la tabla 5 se puede observar la comparación entre este método, y el descrito en la sección anterior, para las redes Lit, APMS y Y2H.

Se puede observar que para las redes LIT e Y2H, los

<sup>2</sup>El análisis de recableado no se realizó para la red Lit\_Reg debido a falta de tiempo.



	$\alpha_1$	$\alpha_2$	$\beta_1$	$\beta_2$
<b>LIT</b>	0.062 +/- 0.014	0.059 +/- 0.005	0.303 +/- 0.065	0.277 +/- 0.016
<b>Y2H</b>	0.015 +/- 0.011	0.025 +/- 0.004	0.180 +/- 0.056	0.2 +/- 0.01
<b>APMS</b>	0.025 +/- 0.014	0.072 +/- 0.003	-----	-----

Cuadro 5 – En esta tabla se presentan las probabilidades  $\alpha$  y  $\beta$  calculadas a partir del método del ajuste lineal (1) y del método del recableo (2).

parámetros calculados mediante ambos métodos son consistentes. Sin embargo, no se pudo aplicar el método para la red APMS. Esto se debe a que en dicha red, la cantidad de enlaces entre proteínas está sobreestimado para la cantidad de nodos presentes en la red. Por lo tanto, al arrojar enlaces esenciales de forma aleatoria se excede automáticamente el número de nodos esenciales, por lo que el procedimiento no se puede utilizar para este tipo de redes.

## 5 De las interacciones esenciales a los módulos esenciales

En el modelo de He la probabilidad de que un nodo sea esencial condicional sobre la esencialidad de otros nodos no vecinos es idéntica a la probabilidad no condicional. Es decir, según el modelo, si dos proteínas no adyacentes son esenciales, entonces el carácter esencial de una es totalmente independiente del hecho de que otra proteína lo sea.

Luego, dados dos nodos no vecinos, la probabilidad de que los dos nodos sean esenciales se reduce al producto de la probabilidad de que el primer nodo sea esencial, por la probabilidad de que el segundo nodo sea esencial. De forma análoga se pueden obtener probabilidades de que ambos nodos sean no esenciales. Más aún, esta probabilidad es la misma incluso si dicho par de nodos tiene vecinos comunes.

Para estudiar este aspecto del modelo, Zotenko propone considerar los pares de nodos dentro de cada red que cumplan con el siguiente criterio: deben ser no vecinos y tener una cantidad mínima de vecinos comunes

(1 o 3 dependiendo de la red considerada y del grado medio de la misma). Luego, se compara el número de pares de nodos del mismo tipo (“esencial” o “no esencial”) observado en cada red con el número de enlaces del mismo tipo que se esperaría con el modelo de He. Para calcular el número esperado de enlaces del mismo tipo, se procede de la siguiente forma.

La probabilidad de tener un enlace del mismo tipo en la red, para dos nodos  $i$  y  $j$  con grados mayor o igual al umbral, está dada por la suma entre el producto de las probabilidades de que ambos sean no esenciales ( $P_N E$ ), con el producto de las probabilidades de que ambos sean esenciales ( $P_E$ ):

$$P_{ij} = P_{E_i} \cdot P_{E_j} + P_{NE_i} \cdot P_{NE_j} \quad (4)$$

Cabe aclarar, además, que la probabilidad individual de cada nodo depende únicamente de su grado y de los parámetros  $\alpha$  y  $\beta$  del modelo.

Consideremos para cada par de nodos  $i$  que satisfaga la condición planteada, la variable aleatoria  $X_i$  dada por  $X = 1$  si ambos nodos son del mismo tipo y  $X = 0$  si no. Ahora consideremos  $N = \sum_i X_i$ , la variable aleatoria “número de pares (que satisfacen la condición y son) del mismo tipo”. Dado que la esperanza es un operador lineal, vemos que

$$E[N] = \sum_i E[X_i] = \sum_i P_i \quad (5)$$

donde  $P_i$  es la probabilidad de que el  $i$ -ésimo par de nodos consista en dos nodos del mismo tipo.

Las probabilidades mencionadas fueron calculadas a partir de la ecuación 1, con valores de  $\alpha$  y  $\beta$  extraídos del ajuste de cada red. Para obtener las incertezas en los números esperados, debidas al error en los parámetros  $\alpha$  y  $\beta$  obtenidos a partir de ajuste lineal, se realizó una pequeña simulación de Montecarlo asignando a  $\alpha$  y  $\beta$  distribuciones normales centradas en los valores nominales calculados y con dispersión igual a su error. Se generaron 1000 pares aleatorios de  $\alpha$  y  $\beta$ , se calculó el número esperado de pares del mismo tipo con cada uno y se calcularon su promedio y su desviación estándar.

Si se procede de esta forma para todas las redes, se puede observar la comparación entre el número de enlaces del mismo tipo observado con el esperado según el modelo de He en la tabla 6.

	# Pares	# Pares del mismo tipo	# Pares esperados del mismo tipo	Z-scores
Lit	9934	5767	5008 +/- 88	8.54
Lit_reg	10777	6187	5790 +/- 147	2.68
APMS	11569	5875	6073 +/- 3048	-0.07
Y2H	23013	15045	14204 +/- 1754	0.48

Cuadro 6 – Para cada red, la cantidad de enlaces entre nodos del mismo tipo observado, y esperado segun el modelo de He.

Como se puede observar en la tabla 6, los resultados observados empíricamente para las redes Lit y Lit-reg no son consistentes con el modelo de He, dado que se encuentran a una gran cantidad de sigmas respecto del valor esperado (Z-score mayor a 1). En cambio, no es posible descartar el modelo a partir de este análisis para las redes APMS y Y2H.

## 6 Conclusiones

Se reprodujeron los resultados de los papers de He[2] y de Zotenko[3]. Por un lado, se estudiaron las hipótesis de He acerca de la relación entre centralidad y esencialidad. Para esto, se calculó la probabilidad de que un nodo aleatorio sea esencial dado que pertenece a un enlace esencial o debido a otro factor externo, y la probabilidad de que un enlace aleatorio sea esencial. Dichas probabilidades se calcularon mediante los dos métodos propuestos por He (una regresión lineal sobre el logaritmo de las probabilidades y un recableado de la red original) y se observó que mientras que el método de la regresión lineal es robusto para todas las redes, el método de recableado falla en redes que sobrestiman el número de enlaces (APMS en nuestro caso).

Por otro lado, se compararon distintas formas de desarmar una red con el objetivo de estudiar los distintos criterios de centralidad. De esta forma, se observó que la remoción de nodos esenciales con el objetivo de

desarmar la componente gigante de la red tiene una eficiencia similar a quitar nodos de forma aleatoria. Por el contrario, la forma más efectiva para desarmar la mayoría de las redes fue con el criterio *betweenness*, el cual ordena la importancia de los nodos según la cantidad de caminos más cortos entre un par de nodos que pasen por el nodo en cuestión.

Finalmente se utilizó el argumento de Zotenko para probar la insuficiencia del modelo de He, el cual no toma en cuenta posibles correlaciones entre la esencialidad de un par de nodos no adyacentes. Para esto, se comparó el número de enlaces del mismo tipo (esencial-esencial o no esencial- no esencial) observados con el número de dichos enlaces, predicho por el modelo de He. Se observó que la cantidad de enlaces esperados es menor que la cantidad de enlaces reales para las redes LIT y LIT REG.

Como observación final, al observar la dirección en la que progresaron los sucesivos modelos que buscan explicar este fenómeno, la historia de la regla de centralidad-letalidad ilustra el surgimiento de la mesoescala como un nivel crítico de la organización biológica"[4].

## Referencias

[1] Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411(6833): 41–42.

[2] He X, Zhang J (2006) Why do hubs tend to be essential in protein networks? *PLoS Genet* 2(6): e88. doi:10.1371/journal.pgen.0020088.

[3] Zotenko E, Mestre J, O’Leary DP, Przytycka TM (2008) Why do hubs in the yeast protein interaction network tend to be essential: re-examining the connection between the network topology and essentiality. *PLoS Comput Biol* 4(8): e1000140. doi:10.1371/journal.pcbi.1000140

[4] Leland H. Hartwell, John J. Hopfield, Stanislas Leibler & Andrew W. Murray (1999) *Nature* volume 402: C47–C52. doi:10.1038/35011540