

UNIVERSIDAD DE BUENOS AIRES

INFORME 2:

Lethality-Centrality Rule

Raúl Barriga
Mariela Celis
Jimmy Masías
Sebastián Pinto

29 de septiembre de 2016

Índice

1. Introducción	2
2. Caracterización	3
3. Regla de Centralidad-Letalidad [Jeong et al., 2000]	4
3.1. Cumplimiento de la regla	5
3.2. Impacto de remoción.	7
3.2.1. Remoción de nodos al azar.	7
4. Hipótesis de interacciones esenciales [He and Zhang, 2006]	10
4.1. Método de Ajuste	11
4.2. Método de simulaciones	13
4.3. Discusión	14

1. Introducción

El modelo clásico de redes aleatorias de Erdős-Rényi asume que un par de nodos están conectados con una probabilidad p , generando redes estadísticamente homogéneas y distribución de grado tipo Poisson [Jeong et al., 2000]. Sin embargo, muchas redes complejas reales (como internet [Faloutsos et al., 1999] o redes metabólicas [Jeong et al., 2000]) muestran comportamientos libre de escala, es decir, se caracterizan por tener pocos nodos altamente conectados *hubs* con nodos poco conectados.

En el caso de redes biológicas, a través de técnicas genéticas, se ha demostrado la existencia de genes indispensables para la sobrevivencia (genes *esenciales*) [Kamath et al., 2003, Winzeler et al., 1999]. Desde entonces se ha buscado una forma de caracterizar la esencialidad de un nodo en una red a través de sus características topológicas.

En Jeong et al. [2000] se trabaja sobre 43 redes metabólicas y reporta una correlación entre los hubs de la red y la esencialidad del nodo desde el punto de vista biológico. A esta correlación la llamaron *regla de Centralidad-Letalidad* (Centrality-Lethality rule) y posteriormente este trabajo ha sido reportado numerosas veces por otros autores. La principal crítica que ha recibido esta correlación es al ser considerada muchas veces como razón causal, es por ello que en He and Zhang [2006] se plantea una aproximación distinta al problema, pero con consecuencias equivalentes. En He and Zhang [2006] se establece que la esencialidad de un nodo (i.e. proteína) se debe a la participación de este en una interacción/proceso (conexiones) esencial, luego debido a que los hubs tienen una alta conectividad con otros nodos, es más probable que una de sus interacciones sea esencial.

El presente estudio es una revisión del trabajo de Zotenko et al. [2008] en que se estudia las debilidades de las hipótesis de Jeong y He. Para ello utilizamos cuatro redes proteicas de levadura (*S. cerevisiae*): Obtenida a partir de Affinity-Purification/Mass-Spectrometry (AP-MS) [Collins et al., 2007], de interacciones binarias yeast two-hybrid (Y2H) [Yu et al., 2008] y dos curadas de literatura [Reguly et al., 2006] (**FALTA UNA CITA AQUI**). Los datos para construir las redes fueron obtenidos de la *Yeast Interactome Database*.

2. Caracterización

Siguiendo el trabajo de Zotenko et al. [2008] tomamos cuatro redes de interacción de levadura (*Saccharomyces cerevisiae*) disponibles en el Yeast Interactome Database¹. Estas redes fueron obtenidas a partir de diferentes técnicas: a través de Affinity-Purification/Mass-Spectrometry (AP-MS) [Collins et al., 2007]; de interacciones binarias Yeast Two-Hybrid (Y2H) [Yu et al., 2008] y dos curadas de literatura [Reguly et al., 2006] (**FALTA UNA CITA AQUI**). La figura 1 muestra una representación gráfica de las cuatro redes consideradas.

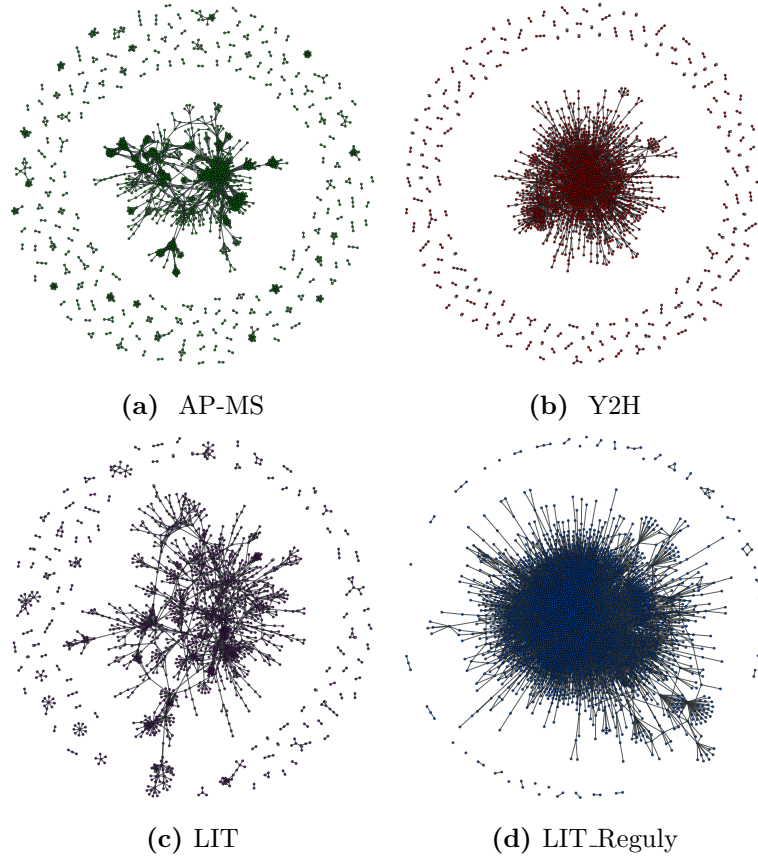


Figura 1: Esquematización de los grafos de cada red estudiada.

La tabla I muestra un resumen de los observables estadísticos despren-

¹<http://interactome.dfci.harvard.edu/S-cerevisiae/>

dados de las propiedades topológicas de las redes. Por otro lado, la tabla II muestra la cobertura entre cada par de redes.

Es importante notar que debido a las diversas técnicas de construcción de las redes, estas presentan diferencias significativas entre sus observables, como por ejemplo, no es extraño que la red AP-MS sea la que presenta mayor clusterización y grado medio debido a que sus enlaces están formados a partir de co-pertenencia a complejos proteicos, mientras que, por ejemplo, Y2H da cuenta de interacciones binarias/físicas entre las proteínas. Además a partir de la tabla II se puede ver que las proteínas (nodos) en la red LIT, están totalmente cubiertas por la red LIT_Reguly, sin embargo esta última sólo cubre un 5 % de los enlaces de la red LIT.

Tabla I: Observables para las cuatro redes de interacción proteica de levadura.

Observables	AP-MS	LIT	Y2H	LIT_Reguly
Nº nodos N	1622	1536	2018	3307
Nº enlaces L	9070	2925	2930	11334
Densidad	0.0068	0.0024	0.0014	0.0020
Diametro	15	19	14	12
Grado k				
medio $\langle k \rangle$	11.18	3.80	2.93	6.85
maximo $\max(\{k\})$	127	40	91	318
minimo $\min(\{k\})$	1	1	1	0
Coefficiente de Clusterización				
medio/local $\langle C \rangle$	0.0710	0.4556	0.0970	0.3583
triangular/global C_{Δ}	0.6185	0.3461	0.0236	0.1241

Tabla II: Observables para las tres redes de interacción proteica de levadura.

AP-MS	0.44	0.57	0.89
0.35	Y2H	0.36	0.66
0.60	0.48	LIT	1.00
0.43	0.40	0.46	LIT_Reguly

3. Regla de Centralidad-Letalidad [Jeong et al., 2000]

En el trabajo de de Jeong et al. [2000] se establece una correlación entre esencialidad y los hubs de una red metabólica. Esta correlación ha sido llamada *Regla de Centralidad-Letalidad* y ha sido reportada en otros trabajos citepyu2004,winzeler1999,wuchty2002. En Jeong et al. [2000] se explora la posibilidad de estimar la esencialidad de un nodo a partir de la centralidad

por grado. En esta sección veremos que si bien si existe una correlación entre esencialidad y los hubs, el comportamiento de estos ultimos (e incluso nodos centrales por otras medidas) no se corresponde con el de un nodo esencial.

3.1. Cumplimiento de la regla

Utilizando una lista de 1155 proteínas esenciales de levadura, se busca la relación entre esencialidad y alto grado de un nodo, donde *alto grado (hub)* debe ser definido. La figura 2 muestra esta relación: en el eje horizontal el porcentaje de nodos que son considerados hubs, y en el eje vertical la fracción de nodos esenciales que además son hubs. De la figura se puede ver que el caso de la red Y2H los nodos esenciales están repartidos homogeneamente en los distintos grados, mientras que el resto de las redes diluye a los esenciales con el aumento del umbral que define a los hubs, es decir, los nodos esenciales de estas redes se encuentran, principalmente, entre los hubs de mayor grado.

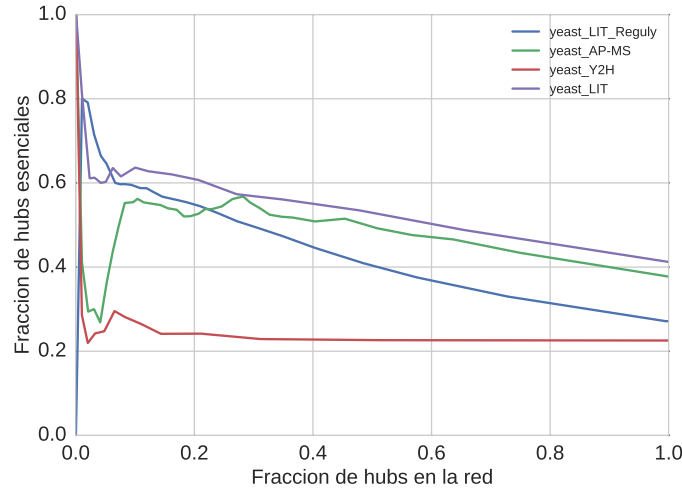


Figura 2: Fracción de hubs esenciales en función de fracción de hubs en relación al total de nodos de la red.

La tabla III muestra los índices de correlación de Kendall (τ) y de Spearman (ρ) de esencialidad con fracción de hubs en la red. Se puede apreciar que incluso para diferentes umbrales de definición de hub, existe una significativa correlación (tanto de Kendall como de Spearman) entre estas dos características; Así también se puede ver que debido a las distintas conectividades de la redes, los grados, a partir de los cuales definimos hub, son

muy distintos para cada red.

Tabla III: Indices de correlación de Rendall (τ) y Spearman (ρ), con sus respectivos p_{valor} , para relación entre esencialidad y grado de un nodo para dos definiciones porcentuales de hubs respecto al total de la red ($\sim 10\%$ y $\sim 20\%$ de los nodos son hubs).

	%10 de corte		
	τ	ρ	k_{cut}
AP-MS	0.62(6.5e-13)	0.70(1.3e-10)	27
LIT	0.89(1.4e-08)	0.94(2.4e-10)	9
Y2H	0.72(4.3e-08)	0.79(3.8e-07)	6
LIT_Reguly	0.84(2.3e-19)	0.91(5.0e-22)	16
	%20 de corte		
	τ	ρ	k_{cut}
AP-MS	0.72(2.7e-19)	0.81(1.4e-17)	18
LIT	0.92(2.7e-10)	0.96(9.6e-14)	6
Y2H	0.75(2.5e-09)	0.84(8.2e-09)	4
LIT_Reguly	0.88(4.9e-23)	0.94(2.2e-28)	10

3.2. Impacto de remoción.

En esta sección estudiamos el impacto de remover proteínas que son centrales bajo diferentes criterios y lo comparamos con el impacto de remover las proteínas catalogadas como esenciales. Según Jeong et al. [2000], un nodo con alto grado en la red de interacciones tiene más probabilidad de ser esencial que la esperada por azar, debido a que actúa de intermediario entre nodos de menor grado. Sin embargo, cabe la posibilidad de que no sea el número de vecinos inmediatos lo que determina la importancia de una proteína, sino algún otra propiedad topológica más global.

En la figura 3 estudiamos cómo se modifica el tamaño relativo (al tamaño de la componente de la red original) de la componente más grande al remover los nodos catalogados como centrales según diferentes criterios, para las cuatro redes estudiadas. Comparamos estos resultados con el caso de remover las proteínas esenciales, y con remover nodos al azar sin más criterio. Los criterios utilizados fueron: grado del nodo, centralidad de betweenness, centralidad de Bonacich, y subgraph centrality.

De la figura 3 concluimos que el impacto de remoción de nodos centrales es mayor que la remoción de los nodos esenciales, que en algunos casos es equivalente a remover nodos al azar. Esto daría la noción que la esencialidad de una proteína no está ligada exclusivamente a su rol en la red de interacciones.

Cabe destacar que el proceso de remoción consistió en calcular la centralidad de cada nodo en la red original, y removerlos de mayor a menor sin volver a realizar el cálculo. Esto implica que removimos los nodos que son centrales en la red original. Si se decide remover el nodo con mayor centralidad luego de realizar una remoción, el impacto de la misma es mayor que el criterio que utilizamos, sin embargo le daríamos el rol de central a un nodo que quizás no lo cumple en la red original. A modo de criterio alternativo, incluimos un gráfico análogo a 3 en la figura 4, donde se puede observar que el impacto de remoción es mayor.

3.2.1. Remoción de nodos al azar.

En las figuras 3 y 4, la remoción de nodos al azar no siguió ningún criterio particular. En esta sección estudiaremos el impacto de tomar nodos al azar no esenciales y cuya distribución de grado sea equivalente a la de los nodos esenciales presentes en la red. Tomaremos nuevamente el tamaño relativo de la componente más grande como impacto de remoción. Cabe recordar que un tamaño relativo menor implica un impacto de remoción mayor.

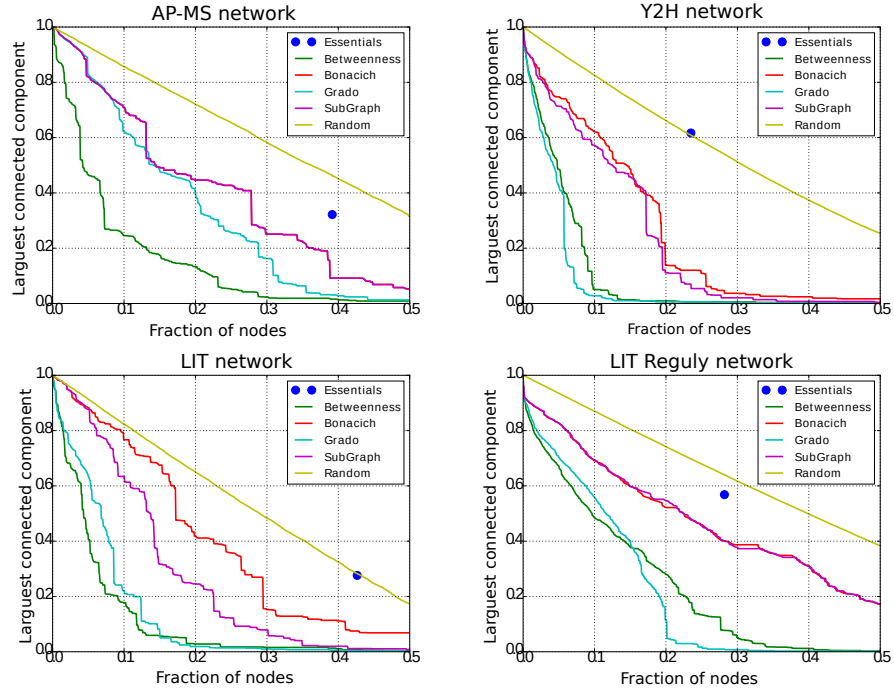


Figura 3: Tamaño relativo de la componente conectada más grande a medida que se remueven los nodos ordenados con diferentes criterios de centralidad. Se incluye el impacto de remover las proteínas esenciales y remover nodos al azar. En todos los casos se puede observar que el impacto de remoción (caída del tamaño del fragmento conectado) es mucho más alto al utilizar un criterio de centralidad en comparación con la remoción de las esenciales, que en algunos casos, prácticamente coincide con la remoción al azar.

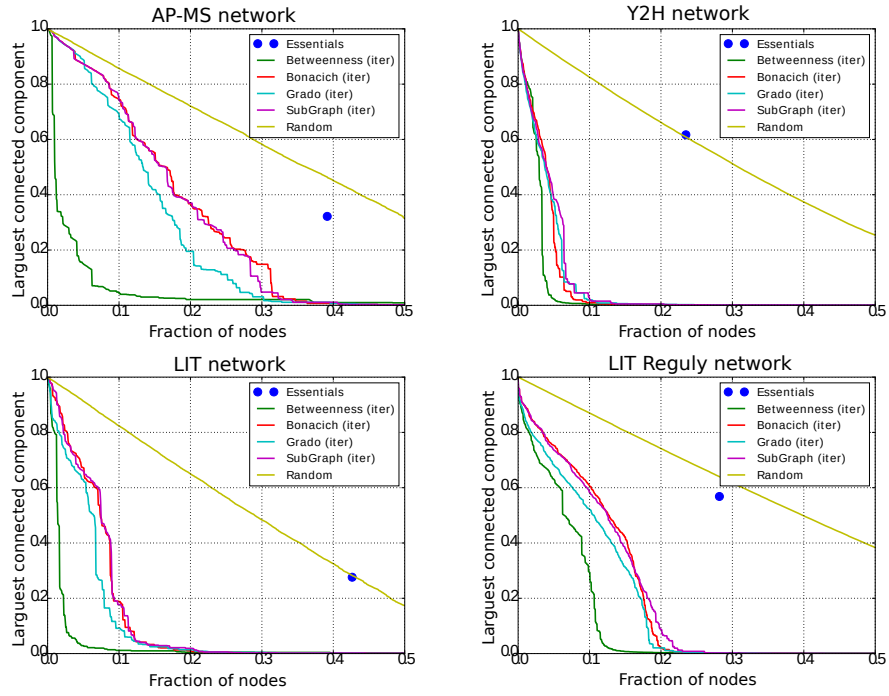


Figura 4: Tamaño relativo de la componente conectada más grande a medida que se remueven los nodos ordenados con diferentes criterios de centralidad, en forma iterativa, es decir, se remueve el nodo más central en la configuración actual de la red. Se puede observar que el impacto de remoción es mayor que el de la figure 3.

Red	Tr(Esenc.)	Tr(No Esenc.)	Nodos removidos
AP-MS	0.32	$0,38 \pm 0,02$	615
LIT	0.276	$0,417 \pm 0,003$	636
Y2H	0.62	$0,61 \pm 0,01$	459
LIT-Reguly	0.568	$0,536 \pm 0,005$	903

Tabla IV: Tamaño relativo de la componente conectada más grande al remover proteínas esenciales, y nodos no esenciales con distribución de grado equivalente.

En la tabla IV, comparamos el impacto de remover los nodos esenciales con la remoción de nodos no esenciales, con distribución de grado equivalente. El criterio tomado consistió en, dado un nodo esencial, tomar uno no esencial con el grado más cercano posible.

De la tabla podemos concluir que en las redes Y2H y LIT-Reguly, el impacto de remoción es prácticamente el mismo que remover nodos no esenciales con la misma distribución de grado, e incluso el impacto es levemente mayor para la segunda red. Sin embargo, para las redes restantes, y con mayor notoriedad en la red LIT, la remoción de las proteínas esenciales es más dañina que la remoción de nodos al azar. Si observamos la figura 3 para la red LIT, la remoción de nodos al azar sin ningún criterio es prácticamente igual a remover los esenciales. Cabe destacar que en la construcción de la figura 3 al tomar nodos al azar se incluyeron a las proteínas esenciales, lo cual, la tabla podría indicar que en esta red algunas proteínas esenciales cumplen algún rol de centralidad.

Del estudio de impacto de remoción concluimos entonces, que ciertas proteínas no son esenciales exclusivamente por su rol topológico en la red, debido a que remover aquellos nodos que cumplen cierto grado de centralidad es siempre más dañino para la red que la remoción de las catalogadas como esenciales. Incluso en algunas de las redes, el impacto de remover nodos al azar es equivalente a la remoción de las esenciales.

4. Hipótesis de interacciones esenciales [He and Zhang, 2006]

Como hemos mostrado en la sección 3 la correlación vista por Jeong es corroborada, sin embargo su hipótesis de esencialidad debido centralidad por grado no responde al comportamiento real de las redes. He y sus colegas plantean una hipótesis alternativa: La esencialidad se debe a la participa-

ción en interacciones esenciales y no esencialidad *unitaria* como pensaba Jeong. Bajo esta hipótesis la regla de Centralidad-Letalidad se explica debido a que los hubs, al tener mayor cantidad de interacciones, tienen mayor probabilidad de pertenecer a una interacción esencial.

El modelo de He considera dos condiciones de esencialidad: La proteína participa en una interacción esencial representada en la red con una probabilidad α ó la proteína es esencial debido a algún otro factor que la red no es capaz de mostrar. Así la probabilidad de que un nodo de grado k no participe en una interacción esencial es $(1 - \alpha)(1 - \alpha) \cdots (1 - \alpha) = (1 - \alpha)^k$ y la probabilidad de no ser esencial debido a otra razón es $(1 - \beta)$. Luego, la probabilidad de no ser esencial es

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

finalmente, la probabilidad de esencialidad es

$$P_E(k) = 1 - (1 - \beta)(1 - \alpha)^k \quad (2)$$

A continuación evaluaremos la calidad de esta hipótesis a través de estimaciones de las probabilidades por dos métodos descritos en el trabajo de He y colegas. Cabe señalar que, para lo que resta del análisis, la red AP-MS no será considerada, debido a que He and Zhang [2006] hace referencia redes construidas a partir de asociación de complejos proteicos podrían presentar comportamientos anómalos.

4.1. Método de Ajuste

Tomando el logaritmo a la ec. 1 se tiene

$$\ln(1 - P_E(k)) = k \ln(1 - \alpha) + \ln(1 - \beta) \quad (3)$$

es decir, una relación lineal entre el grado y la probabilidad de no-esencialidad. He and Zhang [2006] menciona que, debido a que a mayor grado k es más difícil encontrar nodos con grado k , por lo que para sus análisis utiliza un umbral superior de $k = 10$. En este trabajo se ha considerado lo mismo. La figura 5 muestra los resultados del ajuste lineal para nuestras redes. Debido al comportamiento irregular de la red Y2H (ver figura 5c) esta fue excluida del análisis.

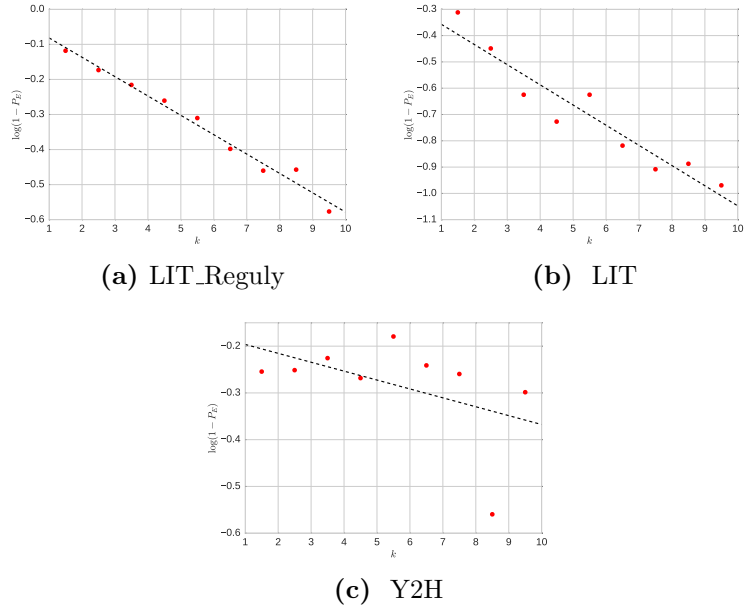


Figura 5: Ajustes lineales de las probabilidades de no-esencialidad y grado (ver ec. 1. (a) $\alpha = 0,53 \pm 0,002$ y $\beta = 0,26 \pm 0,017$; (b) $\alpha = 0,073 \pm 0,008$ y $\beta = 0,224 \pm 0,043$. Debido al comportamiento irregular, Y2H ha sido excluida del análisis.

4.2. Método de simulaciones

Es difícil identificar interacciones esenciales (PPIs) experimentalmente en la escala genómica, dado que esta identificación requiere la demostración de que romper el enlace entre proteínas esenciales sin afectar otros aspectos de las funciones proteicas causa letalidad o infertilidad.

Aquí usamos un método computacional para evaluar la prevalecencia de enlaces esenciales PPIs y la contribución de PPIs esenciales a la esencialidad de genes al nivel genómico.

Nuestro análisis se enfoca en las redes *LIT*, *LIT-Reguly*, y *Y2H*, excluyendo a la red *AP-MS*, como se hizo en la sección anterior.

Como se mencionó antes, dos proteínas que forman un enlace esencial PPI deben ser esenciales. Por el contrario, las interacciones entre proteínas esenciales (IBEPs, *Interaction Between Essential Proteins*, por sus siglas en inglés) pueden o no ser esenciales, dado que la esencialidad de una proteína puede deberse a otros factores además de las PPIs. Esta característica nos permite estimar el número de PPIs esenciales en una red, dado que el número de IBEPs crece con el número de PPIs esenciales.

Dado el número total de interacciones IBEPs N_{ie} para cada red, generamos una red de control haciendo un recableado de los enlaces, manteniendo la distribución de grado $P(k)$ para cada nodo. Repitiendo este procedimiento 5000 veces (1000 para la red *Y2H*), obtenemos la distribución del número (n_{ie}) de enlaces esenciales (IBEPs) en redes recableadas al azar. En todos los casos, el máximo valor de la distribución no supera el caso de la red real; es decir, $\max(n_{ie}) < N_{ie}$ siempre. Este exceso del caso real también se observa en otros casos de PPIs de levadura y en PPIs de nematodos **CITAR 14 y 15, ver p. 2**.

Siguiendo el método de He and Zhang [2006], determinamos la fracción de interacciones esenciales PPI como $\alpha = (N_{ie} - \langle n_{ie} \rangle) / (N_{nod})$, siendo N_{tot} el número total de nodos de la red. Los valores para las diferentes redes se muestran en la tabla V.

4.3. Discusión

Tabla V: Resumen de probabilidades estimadas para cada red a través de método de ajuste y simulaciones.

	Simulación		Ajuste	
	α	β	α	β
LIT_Reguly			0.053 ± 0.002	0.026 ± 0.017
LIT			0.073 ± 0.008	0.244 ± 0.043
Y2H			—	—

Tabla VI: Cantidad de pares totales o de igual característica (ambos esenciales o ambos no esenciales) en comparación al valor estimado desprendido de la hipótesis de He and Zhang [2006].

	Pares Totales	Pares del mismo tipo	Pares esperados del mismo tipo	
			Simulación	Ajuste
LIT_Reguly	10777	6187		5716
LIT	1858	1059		963
Y2H	2258	1493		

Referencias

- Sean R Collins, Patrick Kemmeren, Xue-Chu Zhao, Jack F Greenblatt, Forrest Spencer, Frank CP Holstege, Jonathan S Weissman, and Nevan J Krogan. Toward a comprehensive atlas of the physical interactome of *saccharomyces cerevisiae*. *Molecular & Cellular Proteomics*, 6(3):439–450, 2007.
- Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM computer communication review*, volume 29, pages 251–262. ACM, 1999.
- Xionglei He and Jianzhi Zhang. Why do hubs tend to be essential in protein networks? *PLoS Genet*, 2(6):e88, 2006.
- Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.
- Ravi S Kamath, Andrew G Fraser, Yan Dong, Gino Poulin, Richard Durbin, Monica Gotta, Alexander Kanapin, Nathalie Le Bot, Sergio Moreno, Marc

Sohrmann, et al. Systematic functional analysis of the *caenorhabditis elegans* genome using *rnai*. *Nature*, 421(6920):231–237, 2003.

Teresa Reguly, Ashton Breitkreutz, Lorrie Boucher, Bobby-Joe Breitkreutz, Gary C Hon, Chad L Myers, Ainslie Parsons, Helena Friesen, Rose Oughtred, Amy Tong, et al. Comprehensive curation and analysis of global interaction networks in *saccharomyces cerevisiae*. *Journal of biology*, 5(4):1, 2006.

Elizabeth A Winzeler, Daniel D Shoemaker, Anna Astromoff, Hong Liang, Keith Anderson, Bruno Andre, Rhonda Bangham, Rocio Benito, Jef D Boeke, Howard Bussey, et al. Functional characterization of the *s. cerevisiae* genome by gene deletion and parallel analysis. *science*, 285(5429):901–906, 1999.

Haiyuan Yu, Pascal Braun, Muhammed A Yildirim, Irma Lemmens, Kavitha Venkatesan, Julie Sahalie, Tomoko Hirozane-Kishikawa, Fana Gebreab, Na Li, Nicolas Simonis, et al. High-quality binary protein interaction map of the yeast interactome network. *Science*, 322(5898):104–110, 2008.

Elena Zotenko, Julian Mestre, Dianne P O’Leary, and Teresa M Przytycka. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Comput Biol*, 4(8):e1000140, 2008.