

Centralidad vs. Letalidad

Centrality/Lethality rule

- Jeong 2001 Nature, *“Lethality and centrality in protein networks”*
- He 2006 PLOS Genetics Bio *“Why do hubs tend to be essential in protein networks”*
- Zotenko 2008 PLOS Comp Bio, *“Why do hubs in yeast protein interaction network tend to be essential”*

Escencialidad y topología

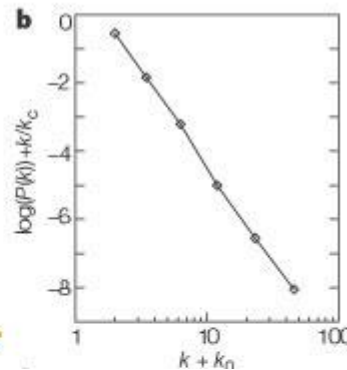
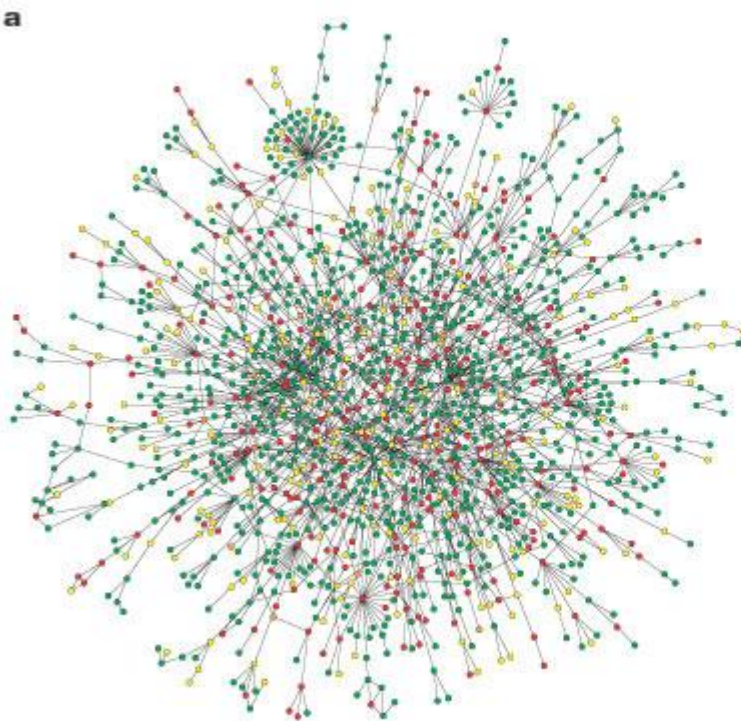
según Jeong



Puede la **esencialidad** de una proteína explicarse en términos de su **ubicación** en una red de interacción de proteínas? (i.e. centralidad en red <-> importancia biológica ?)

Jeong 2001, Nature, “Lethality and centrality in protein networks”

“The most highly connected proteins in the cell are the most important for its survival”



$$p_x = a(k + k_{\text{sat}})^{-\gamma} \exp\left(-\frac{k}{k_{\text{cut}}}\right)$$

Cut-off exponencial $k_c \sim 20$

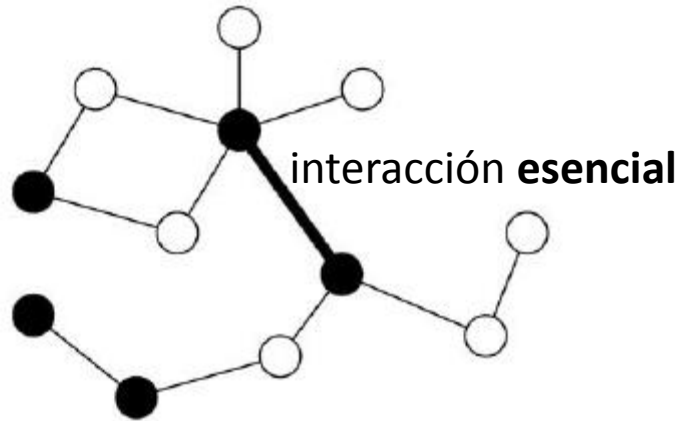
Corrección bajo grado $k_0=1$

- ~21% de $k=5$ son esenciales
- ~60% de proteínas $k=20$ lo son.

Jeong: sugiere que **hubs** son **esenciales** **porque** **mantienen la conectividad** de la red

Esencialidad y topología

según He



- Algunas **proteínas son esenciales** porque participan en **interacciones esenciales** distribuidas uniformemente en la red
- Hubs tienden a ser esenciales simplemente porque están involucrados en mayor número de interacciones

Modelo:

1. Distribuir sobre la red **interacciones esenciales** uniformemente al azar con prob α
2. Distribuir sobre la red **proteínas esenciales** uniformemente al azar con prob β

α : prob. unif. de que una interacción de la red sea esencial

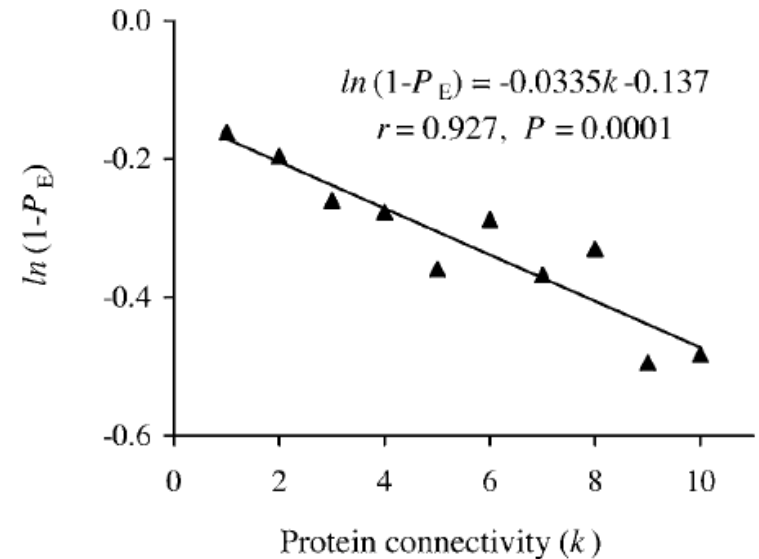
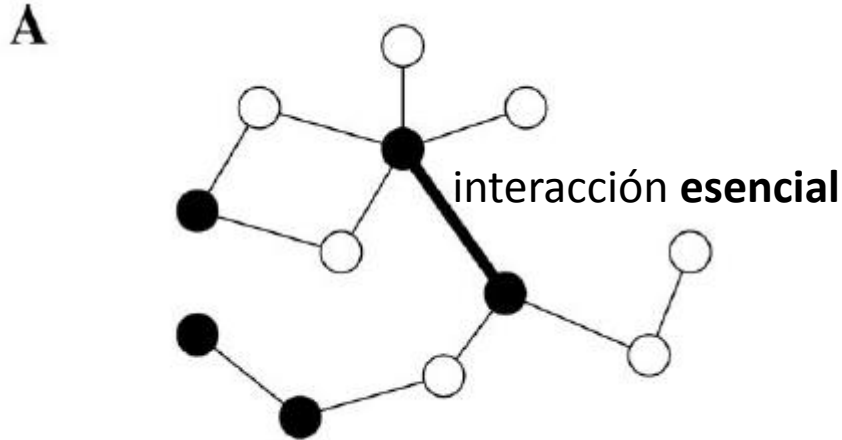
β : prob. de que un nodo sea esencial por otra causa

Prob de que un nodo sea esencial:

$$P_E = 1 - \underbrace{(1 - \beta)(1 - \alpha)^k}_{\text{prob de que un nodo no sea esencial}}$$

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

Metodo 1 para estimar parametros



Modelo

α : prob. de que una interacción sea esencial

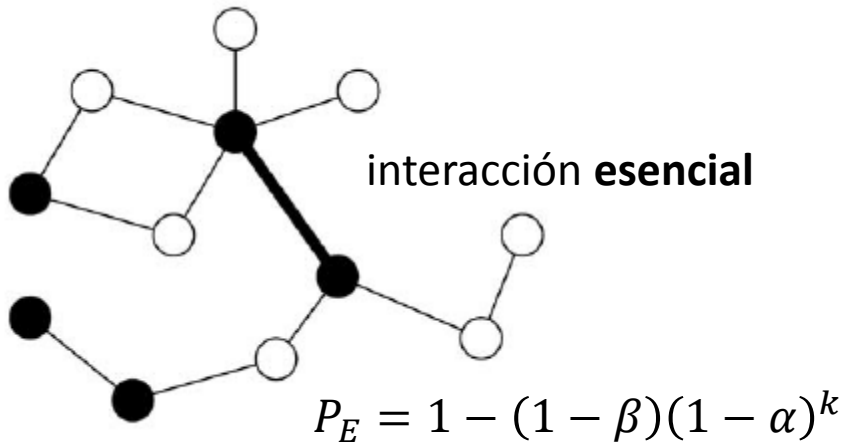
β : prob. de que un nodo sea esencial por otra causa

Prob de que un nodo sea esencial:

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

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

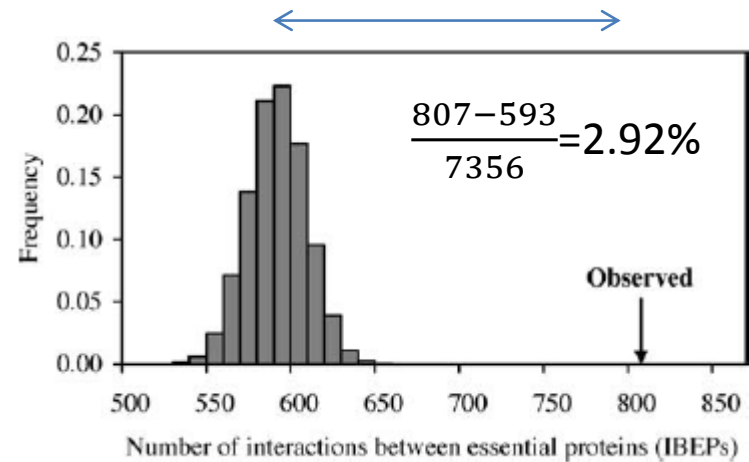
Metodo 2 para estimar parametros



α : prob. de que una interacción sea esencial
 β : prob. de que un nodo sea esencial por otra causa

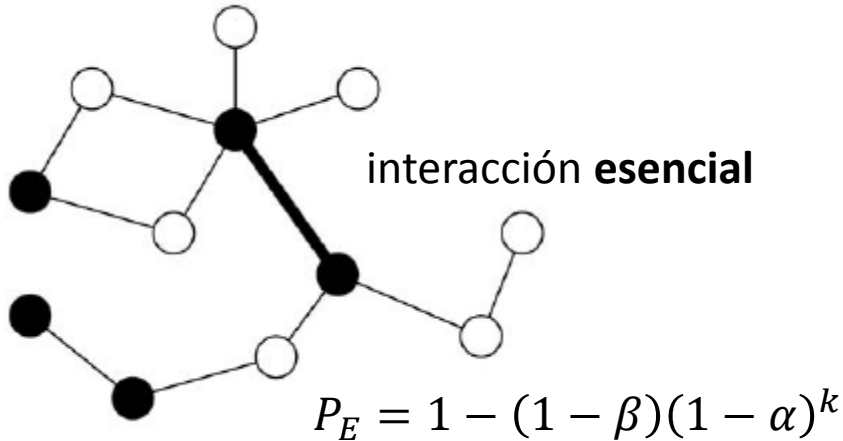
- proteínas linkeadas por ie's deben ser esenciales
- proteínas esenciales pueden estar ligadas por links no esenciales

- PIN levadura 4126 Nodos, 7356 enlaces
- 836 prot esenciales, 807 IBEPs
- En redes rnd (modelo configuracional): $\langle \text{IBEP} \rangle = 592$



Asumiendo que la diferencia en IBEPs se debe a falta de links esenciales ...el exceso de IBEPs **sirve para estimar α**

Metodo 2 para estimar parametros



α : prob. de que una interacción sea esencial

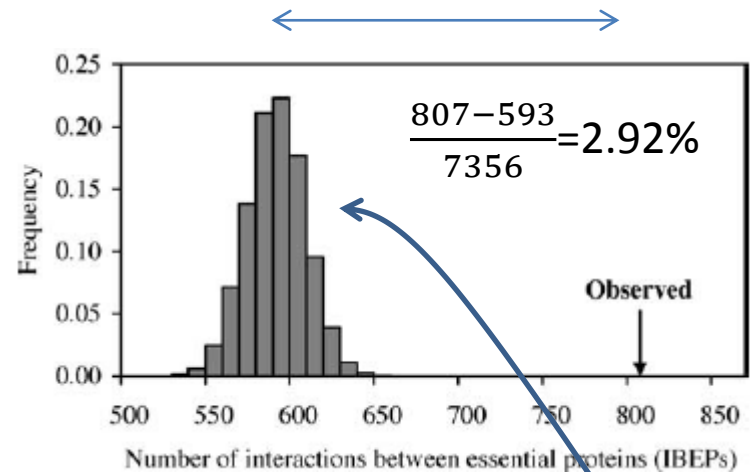
β : prob. de que un nodo sea esencial por otra causa

2) Para estimar β :

- I. elimino etiquetado *esencial*
- II. Asignar al azar 807-m **interacciones esenciales**. m es un nro aleatorio de
- III. Definir nodos esenciales a partir de esas interacciones
- IV. Marcar proteínas esenciales adicionales eligiendo nodos al azar hasta alcanzar las M proteínas esenciales. Estimar β como la fracción de nodos adicionada (nota: un nodo puede haber salido esencial por II y por III)
- V. Repetir 10000

1) Red recableada

exceso de i.e. sirve para estimar α



Entonces....?

hubs son **esenciales** porque **mantienen la conectividad** de la red !

La conectividad de la red (topología global) no pincha ni corta! Lo **esencial** son las **interacciones, papá!**

Ni lo uno, ni lo otro... Lo que puede ser **esencial** ocurre en otra escala (mesosopica). Son los **complejos de proteínas** que llevan adelante funciones biológicas

Chiques...Lo **esencial** es invisible a los ojos



Esencialidad y topología según Zotenko



OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality

Elena Zotenko¹, Julian Mestre¹, Dianne P. O'Leary^{2,3}, Teresa M. Przytycka^{4*}

1 Max-Planck Institute for Informatics, Saarbruecken, Germany, **2** Department of Computer Science, University of Maryland, College Park, Maryland, United States of America, **3** Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland, United States of America, **4** National Center of Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, United States of America

Abstract

The *centrality-lethality rule*, which notes that high-degree nodes in a protein interaction network tend to correspond to proteins that are essential, suggests that the topological prominence of a protein in a protein interaction network may be a good predictor of its biological importance. Even though the correlation between degree and essentiality was confirmed by many independent studies, the reason for this correlation remains illusive. Several hypotheses about putative connections between essentiality of hubs and the topology of protein-protein interaction networks have been proposed, but as we demonstrate, these explanations are not supported by the properties of protein interaction networks. To identify the main topological determinant of essentiality and to provide a biological explanation for the connection between the network topology and essentiality, we performed a rigorous analysis of six variants of the genomewide protein interaction network for *Saccharomyces cerevisiae* obtained using different techniques. We demonstrated that the majority of hubs are essential due to their involvement in Essential Complex Biological Modules, a group of densely connected proteins with shared biological function that are enriched in essential proteins. Moreover, we rejected two previously proposed explanations for the centrality-lethality rule, one relating the essentiality of hubs to their role in the overall network connectivity and another relying on the recently published essential protein interactions model.

Datos de partida



- **DIP-CORE** DIP database
- **LC** interacciones extraídas a partir de 30000 abstracts que reportan experimentos de escala chica. Mayor calidad, problema de sesgo de proteínas de estudio (creación de hubs)
- **HC** Dataset de Batada. Combina interacciones robustas relevadas en experimento de baja y alta escala.
- **Y2H**
- Dataset de Ito.
- **TAP-MS** Datos de purificación de complejos
- **Bayesian** Red inferida computacionalmente a partir de datos de expresión, colocalización,, etc

Table 1. Structural properties of the tested protein interaction networks.

	Number of nodes	Number of edges	Average degree	Average clustering coefficient	
DIP CORE	2,316	5,569	4.81	0.30	binaria y complejos
LC	3,224	11,291	7.00	0.36	
HC	2,752	9,097	6.61	0.37	
TAP-MS	1,994	15,819	15.87	0.60	pertenencia a complejos
BAYESIAN	4,135	20,984	10.15	0.26	
Y2H	400	491	2.45	0.09	int binarias

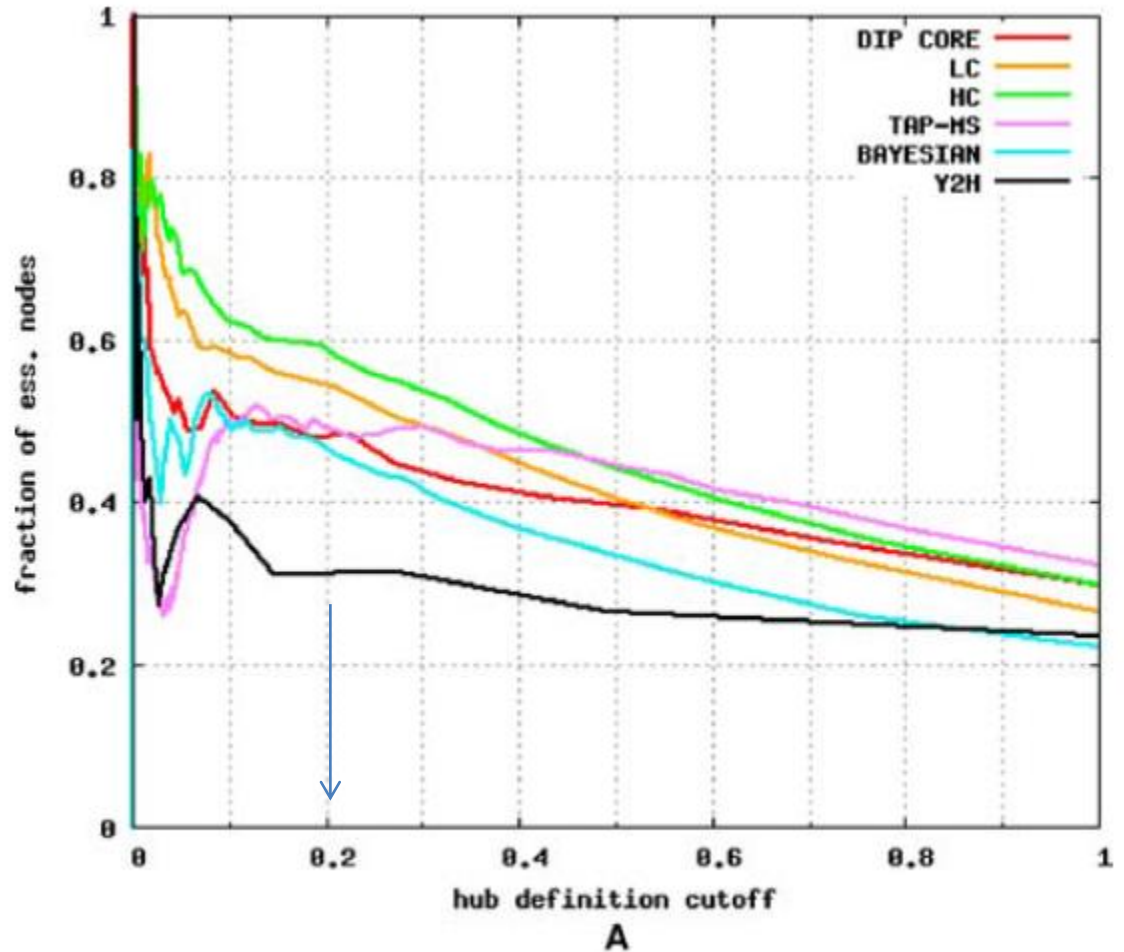
Table 2. Amount of overlap between tested networks.

DIP CORE	0.58	0.62	0.25	0.61	0.02
0.28	LC	0.53	0.26	0.39	0.01
0.38	0.65	HC	0.47	0.47	0.02
0.09	0.18	0.27	TAP-MS	0.36	0.00
0.16	0.21	0.20	0.27	BAYESIAN	0.02
0.26	0.18	0.31	0.10	0.97	Y2H9

La regla de centralidad-letalidad se observa en las 6 redes



1105 genes esenciales *



Impacto de remoción 1

- Efecto de remoción de proteínas esenciales no difiere del de proteínas no-esenciales tomadas al azar

Table 3. Impact of the removal of essential proteins as compared to the removal of an equivalent number of random nonessential proteins with the same degree distribution.

	Essential	Random nonessential
DIP CORE	0.519	0.504±0.007
LC	0.578	0.551±0.010
HC	0.521	0.525±0.005
TAP-MS	0.512	0.512±0.011
BAYESIAN	0.685	0.625±0.006
Y2H	0.410	0.397±0.046

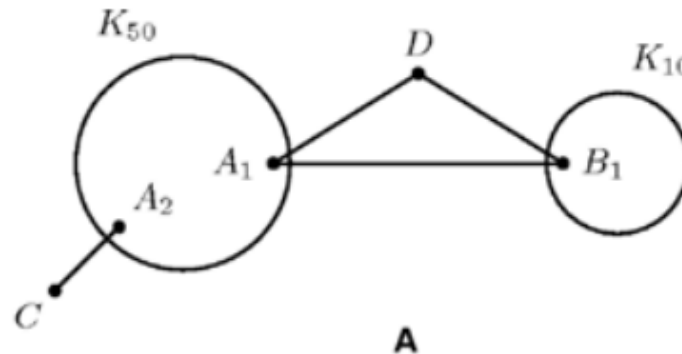
The impact of removal of a set of proteins is measured by the fraction of nodes in the largest connected component. For each network the effect of the removal of essential proteins and the removal of an equivalent number of random nonessential proteins with the same degree is shown.

doi:10.1371/journal.pcbi.1000140.t003

Impacto de remoción 2

Centralidades y vulnerabilidad/conectividad

K50: cjto de 50
 vertices conectados
 completamente
 entre si (*clique*)



K10: cjto de 10
 vertices
 completamente
 entre si (*clique*)

dc: degree centrality

ec: eigenvector centrality

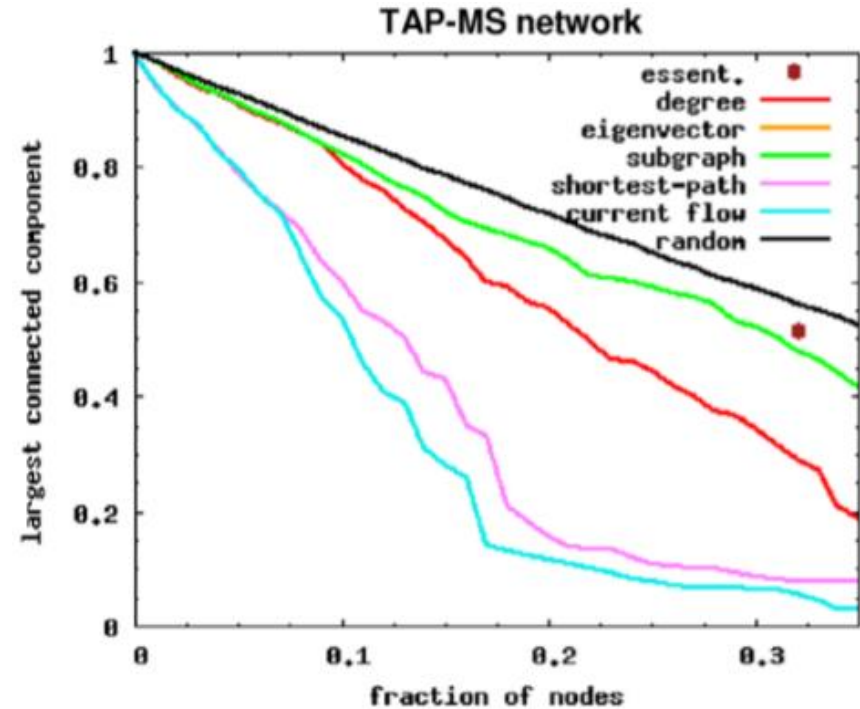
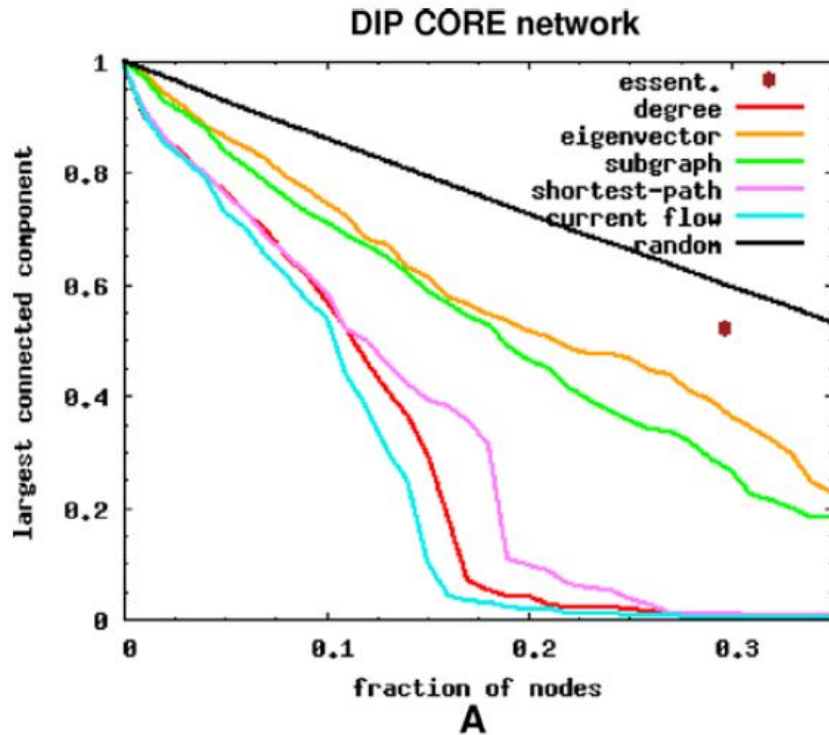
sc: subgraph centrality ,nro de
 ciclos que se originan en un nodo

spbc: short-path betweenness

current-flow centrality: como
 spbc, pero considerando
 multiple paths

dc	ec	sc	spbc	cfc
A1	A1	A1	A1	A1
A2	A2	A2	B1	B1
A3...A50	A3...A50	A3...A50	A2	D
B1	B1	B1	A3...A50, B2...B10, C, D	A2
B2...B10	D	D		B2...B10
D	C	C		A3...A50
C	B2...B10	B2...B10		C

Centralidades y vulnerabilidad/conectividad



Atacar por grado es eficiente en DIP CORE, pero no en TAP-MS

En general medidas locales son menos disruptivas

Hubs esenciales **NO** son particularmente responsables de mantener la conectividad de la red

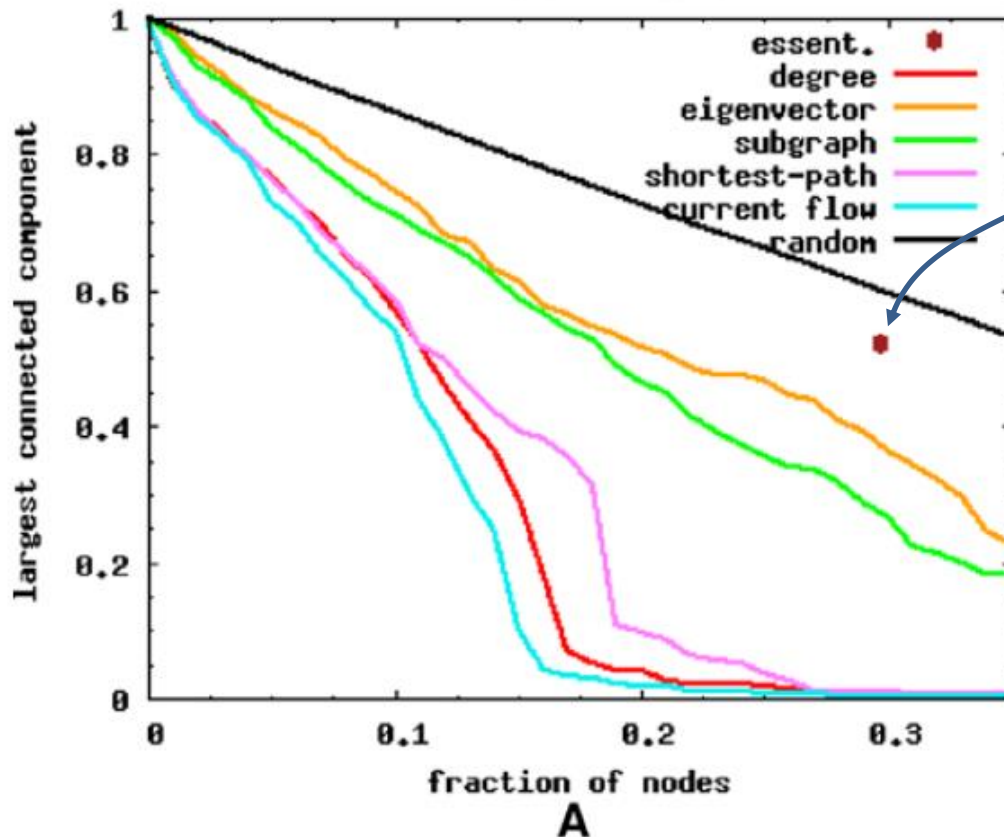
0



1



DIP CORE network



- Retirar genes esenciales de la red es menos disruptivo que remover nodos de acuerdo a cualquiera de los indices estudiados (tiene casi el mismo efecto en el tamaño de la componente gigante que eliminar la misma cantidad de nodos al azar)
- La disrupcion observada al quitar hubs no está relacionada con esencialidad
- **Genes esenciales no parecen cumplir rol de conectividad**

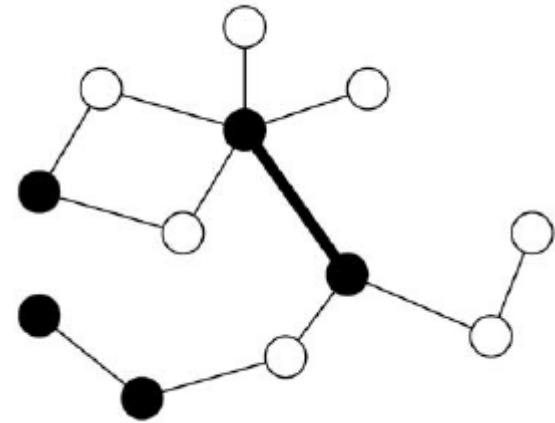
El modelo de interacciones esenciales tiene problemas

Prob de que un nodo sea esencial:

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

Segun el modelo, la probabilidad de que una proteina PE1 sea esencial es independiente de que otra PE2 (no adyacente)lo fuera

Son del mismo tipo (esencial/no-esencial) pares de proteinas, no adyacentes con 3-4 vecinos comunes (i.e. están en un mismo entorno)?



El modelo de interacciones esenciales tiene problemas

0

1+1



Prob de que un nodo sea esencial:

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

Segun el modelo, la probabilidad de que una proteina PE1 sea esencial es independiente de que otra PE2 (no adyacente) lo fuera

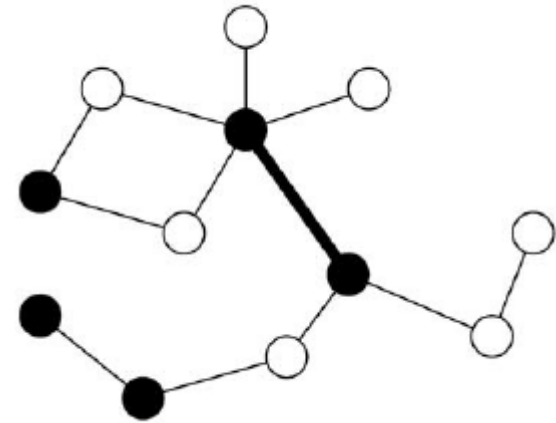


Table 5. Difference between the observed and expected number of pairs where both proteins are either essential or nonessential.

	Total number of pairs	Number of pairs of the same type	Expected number of pairs of the same type		
			Simulation	Line fitting	Weighted line fitting
DIP CORE	1,849	1,135	945 (3.6e-10)	938 (8.6e-12)	938 (8.0e-11)
LC	10,777	6,785	5,691 (6.6e-10)	5,596 (1.1e-15)	5,589 (3.9e-14)
HC	5,907	3,516	3,213 (2.0e-08)	2,997 (2.2e-16)	2,994 (2.2e-16)
Y2H	3,254	2,167	1,976 (9.6e-07)	2,025 (2.6e-04)	2,052 (3.3e-03)

$$P_E(i, j) \neq P_E(i) P_E(j)$$

The total number of pairs refers to the number of nonadjacent protein pairs with three or more common neighbors in the network. (Due to the sparsity of the Y2H network, the statistics are calculated for nonadjacent pairs having one or more neighbors in common.) The nodes in the pair are of "the same type" if they are both essential or both nonessential.

TC02

- Utilizando las redes provistas en la practica
- Reanalizar los datos repitiendo los pasos de Zotenko
 - Caracterizacion de redes (Tabla 1 y Tabla 2)
 - Letalidad Centralidad (Figura 1a)
 - Analisis de vulnerabilidad (Figura 3, considerando: degree, Bonacich, betweenness, subgraph, random)
 - Impacto de remoción (Tabla 3)
 - Falla del modelo de interacciones esenciales (Tabla 5 (alfa y beta a partir de simulacion y fiteo))

Proyecto final



OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality

Elena Zotenko¹, Julian Mestre¹, Dianne P. O'Leary^{2,3}, Teresa M. Przytycka^{4*}

1 Max-Planck Institute for Informatics, Saarbruecken, Germany, **2** Department of Computer Science, University of Maryland, College Park, Maryland, United States of America, **3** Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland, United States of America, **4** National Center of Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, United States of America

Abstract

The *centrality-lethality rule*, which notes that high-degree nodes in a protein interaction network tend to correspond to proteins that are essential, suggests that the topological prominence of a protein in a protein interaction network may be a good predictor of its biological importance. Even though the correlation between degree and essentiality was confirmed by many independent studies, the reason for this correlation remains illusive. Several hypotheses about putative connections between essentiality of hubs and the topology of protein-protein interaction networks have been proposed, but as we demonstrate, these explanations are not supported by the properties of protein interaction networks. To identify the main topological determinant of essentiality and to provide a biological explanation for the connection between the network topology and essentiality, we performed a rigorous analysis of six variants of the genomewide protein interaction network for *Saccharomyces cerevisiae* obtained using different techniques. We demonstrated that the majority of hubs are essential due to their involvement in Essential Complex Biological Modules, a group of densely connected proteins with shared biological function that are enriched in essential proteins. Moreover, we rejected two previously proposed explanations for the centrality-lethality rule, one relating the essentiality of hubs to their role in the overall network connectivity and another relying on the recently published essential protein interactions model.