

# Redes TP02

villagran.azuara.marco

October 2018

## 1 Esencialidad de las interacciones entre proteínas

La verificación experimental de la esencialidad de una proteína en un sistema es un proceso complejo y que no da resultados completamente confiables, ya que la metodología del experimento muy seguramente llega a afectar el resultado al dañar el sistema. En He y Zang (2006)[1] los autores tratan el problema de la determinación de la esencialidad de una proteína planteando la existencia de enlaces (o interacciones entre proteínas) esenciales que unen dichos elementos. La hipótesis elemental del trabajo mencionado consiste en declarar a un nodo (o proteína) central como esencial, por el hecho de tener muchas interacciones, de las cuales un número importante son esenciales. El tener un nodo declarado como esencial por las condiciones ya expuestas relaja la dependencia que tiene la esencialidad de las proteínas con la topología de la red de sus interacciones, esto en comparación con trabajos anteriores, e. g. Jeong y col. (2001)[2].

Teniendo una base de datos de  $I$  interacciones entre  $p$  proteínas y comparándolo con una lista confiable de proteínas esenciales ( $N_E$ ) conocidas se puede obtener el número de interacciones entre  $N_E$ s en una red, llamemos a este número  $I_{PE}$ . Este número puede ser mayor al número de interacciones esenciales reales del sistema,  $I_E$ , ya que no todas las interacciones entre  $N_E$ s son esenciales, pues la esencialidad de una proteína puede ser dada por factores ajenos a una  $I_{PE}$  (véase He y Zang 2006[1]).

Para probar la hipótesis propuesta se generan redes de control re-cableando aleatoriamente (y repitiendo el proceso una cantidad significativa de veces) las conexiones entre proteínas cuidando que cada una de ellas conserve su grado  $k$  de conectividad. En cada re-cableado se puede calcular el número de interacciones entre proteínas esenciales en la red aleatoria, que llamaremos  $RI_{PE}$ . Con este número en mente se define la cantidad  $\alpha$ ,

$$\alpha = \frac{I_{PE} - \langle RI_{PE} \rangle}{I}, \quad (1)$$

que representa el porcentaje de  $I_E$  debida solamente a  $I_{PE}$ <sup>1</sup>. Para tomar en cuenta los factores adicionales que pueden transformar una proteína en esencial se definirá la cantidad  $\beta$ . Para obtener este número generaremos redes aleatorias

---

<sup>1</sup>Suponiendo que siempre  $I_{PE} > RI_{PE}$ .

donde generaremos redes aleatorias en las que se remueve la esencialidad de las proteínas y después se asignan aleatoriamente ( $I_{PE}$  -  $RI_{PE}$ ) enlaces, donde  $RI_{PE}$  se selecciona aleatoriamente de la distribución que generamos al calcular  $\alpha$ . Después se declaran como esenciales los nodos que estén unidos por estas interacciones, la cantidad de nodos marcados de esta forma generalmente será menor al número de nodos que se tienen originalmente. A continuación se agregan aleatoriamente los nodos restantes, lo cual emula las proteínas que son esenciales por factores ajenos. Este proceso puede incluir nuevos enlaces esenciales en la red, con lo cual se puede definir  $\beta$  como el promedio de la fracción de nodos esenciales agregados en cada repetición.

Con  $\alpha$  y  $\beta$  se consideran las dos formas por las que una proteína puede ser esencial, con lo que la probabilidad de que una proteína sea esencial es

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

sin embargo se considerará la probabilidad de que una proteína no sea esencial, i.e.  $1 - P_E$ .

Figure 1: Para la red Yeast AP-MS. Probabilidad de no esencialidad para las proteínas de  $k$ -ésimo grado. Los triángulos indican los resultados directos de la muestra mientras que la línea continua es el resultado del modelo.

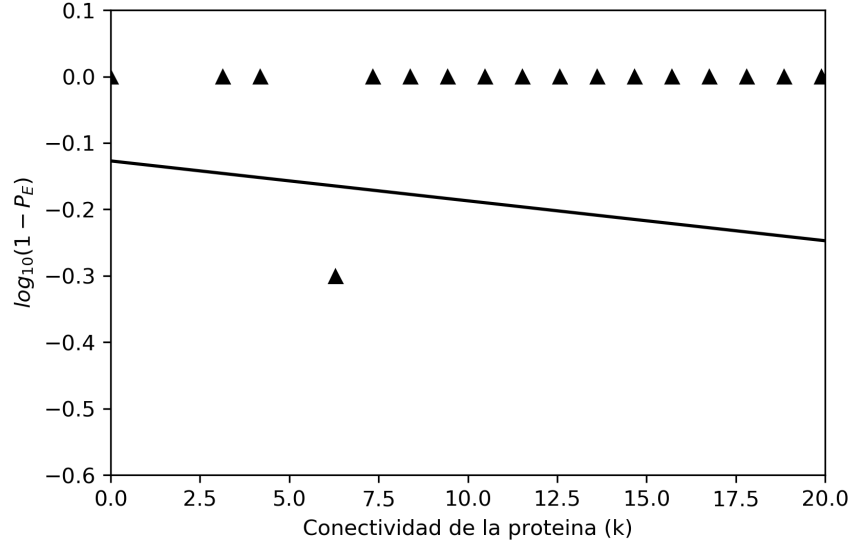


Figure 2: Para la red Yeast Y2H. Probabilidad de no esencialidad para las proteínas de  $k$ -ésimo grado. Los triángulos indican los resultados directos de la muestra mientras que la línea continua es el resultado del modelo.

El resultado de aplicar este análisis a las 3 bases de datos descritas en este trabajo se presentan en las figuras 1, 2 y 3. Adicionalmente los valores numéricos se muestran en la tabla 1.

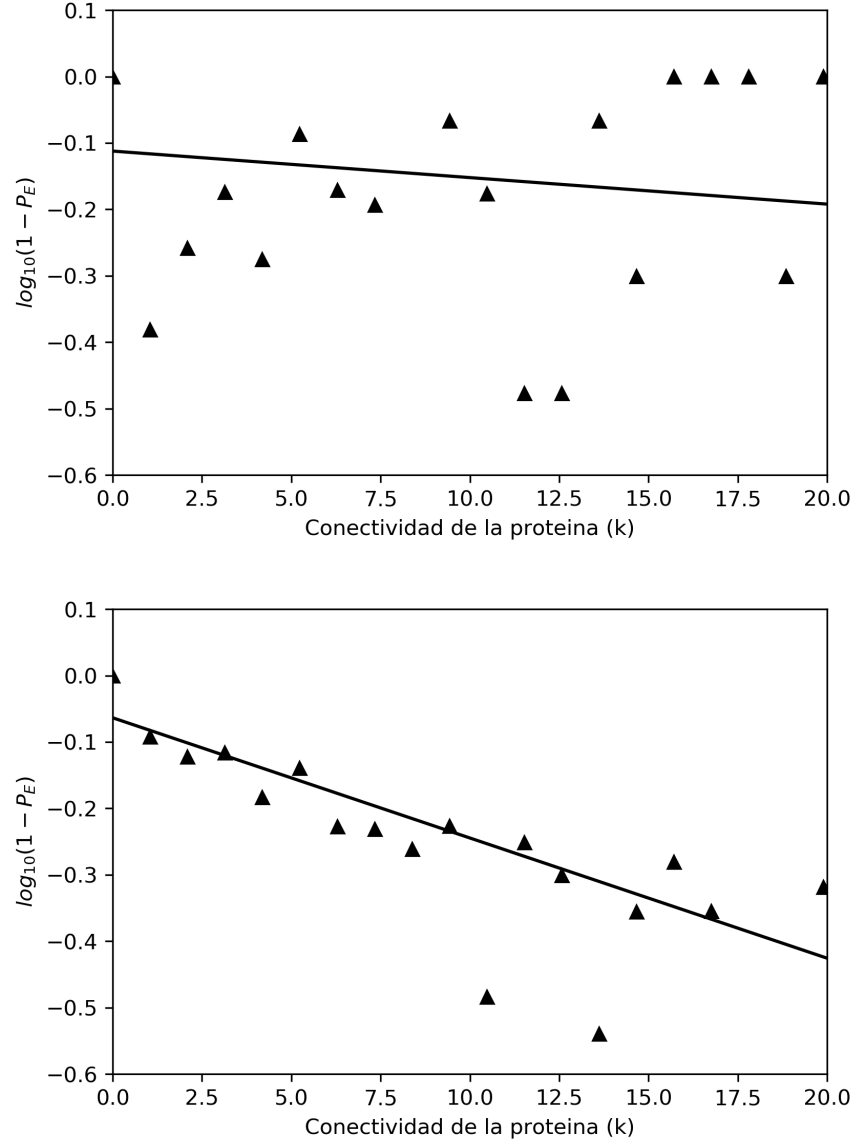


Figure 3: Para la red Yeast Lit. Probabilidad de no esencialidad para las proteínas de  $k$ -ésimo grado. Los triángulos indican los resultados directos de la muestra mientras que la línea continua es el resultado del modelo.

Con el resultado de este experimento se puede apreciar que el tamaño de la muestra favorece el ajuste obtenido con esta metodología. La alta dependencia del resultado con el tamaño de la muestra puede ser un indicador de las fallas de

Datos	Nodos	Nodos Esenciales	Enlaces	Enlaces Esenciales	$\alpha$	$\beta$
yeast_LIT_Reguly	3292	897	11853	6518	0.040	0.14
yeast_ap-ms	72	19	59	16	0.014	0.13
yeast_Y2H	2013	459	2918	546	0.009	0.11

esta metodología y por lo tanto, de las fallas que tiene el pensar que los nodos centrales son esenciales por la cantidad de conexiones que tiene. Adicionalmente es importante mencionar que a pesar de contar con muestras de un tamaño significativo las proteínas con alto grado de conectividad no se muestran, pues de la misma forma que argumentan He y Zang, estos nodos altamente conectados siguen siendo una sub-muestra muy pequeña de proteínas.

## References

- [1] Xionglei He and Jianzhi Zhang. Why do hubs tend to be essential in protein networks? *PLOS Genetics*, 2(6):1–9, 06 2006.
- [2] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411:41 EP –, May 2001.