

Introducción a Redes Complejas en Biología de Sistemas

Trabajo Computacional 1

- 1) Considere las tres redes de interacción de proteínas relevadas para levadura disponibles en la página de la materia. Se trata de: una red de interacciones binarias (yeast_Y2H.txt), de co-pertenencia a complejos proteicos (yeast_AP-MS.txt) y obtenida de literatura (yeast_LIT.txt) obtenidas del Yeast Interactome Database.
 - a. Presente una comparación gráfica de las 3 redes.
 - b. Resuma en una tabla las siguientes características de dichas redes
 - i. El número total de nodos, N
 - ii. El número total de enlaces L , de la red
 - iii. Si se trata de una red dirigida o no-dirigida
 - iv. El grado medio $\langle k \rangle$ ($\langle k_{in} \rangle$, $\langle k_{out} \rangle$ en caso de red dirigida), el grado máximo y mínimo de la red
 - v. La densidad de la red
 - vi. Los coeficientes de clustering $\langle C_i \rangle$ y C_Δ de la red.
 - vii. Diámetro de la red.
 - c. Teniendo en cuenta la naturaleza de las interacciones reportadas, diga si es razonable lo que encuentra para ciertos observables calculados.
 - d. Construya un diagrama de Venn que permita reconocer la cobertura, especificidad y coherencia de las interacciones reportadas por los tres datasets.

- 2) Considere la red social de 62 delfines de Nueva Zelanda (dolphins.txt).
 - a. Examine diferentes opciones de layout para este grafo e identifique la que le resulte más informativa. Justifique su elección detallando las características estructurales de la red que su elección pone en evidencia. Incluya en la representación gráfica de la red información sobre el sexo de los delfines.
 - b. Se trata una red donde prevalece la *homofilia* en la variable género? Para responder
 - i. Considere la distribución nula para la fracción de enlaces que vinculan géneros diferentes, generada a partir de al menos 1000 asignaciones aleatorias de género.
 - ii. A partir de lo obtenido proponga una estimación para el valor y el error de dicha cantidad cuando no existe vínculo entre topología de la red medio y asignación de género. Compare su estimación con el valor medio esperado.
 - iii. Estime la significancia estadística (p-valor) del valor observado en el caso de la red real.
 - c. (*) Identifique alguna metodología basada en observables topológicos para eliminar nodos secuencialmente de la red de manera de dividirla en dos componentes de tamaños comparables en el menor número de pasos. Explique y muestre los resultados obtenidos. Intente cuantificar su estrategia comparándola con lo que se obtendría al eliminar nodos de manera aleatoria.

- 3) Considere la red *as-22july06.gml* creada por Mark Newman que contiene la estructura de los sistemas autónomos de internet relevada a mediados de 2006.
- Encuentre gráficamente la distribución de grado P_k como función de k explorando diferentes alternativas: un bineado lineal o logarítmico, utilizando escalas logarítmicas o lineales en uno o ambos ejes. Discuta que alternativa permite apreciar mejor el carácter libre de escala de dicha distribución.
 - Utilizando funcionalidad de la librería *igraph*, estime el exponente de dicha distribución.

4) Asortatividad

- Considere la red de colaboraciones científicas (*netscience.gml*) y la red de internet (*as-july06.gml*). Analice si nodos de alto grado tienden a conectarse con nodos de alto grado o por el contrario suelen conectarse a nodos de bajo grado? (i.e la red es asortativa o disortativa respecto al grado?). Para ello:
 - Determine, para nodos de grado k , cuánto vale en media el grado de sus vecinos. [hint R: se puede estimar primero el grado medio de los vecinos de cada nodo de la red y luego utilizar *aggregate* sobre esos datos, que permite estimar cantidades sobre subconjuntos determinados de datos de acuerdo a diferentes criterios]
 - Analizar la tendencia observada en un gráfico que consigne dicho valor $k_{nn}(k)$ como función del grado.
 - Asumiendo que $k_{nn}(k) = ak^\mu$, estime el exponente de correlación a partir de realizar una regresión de $\log k_{nn} \sim \log k$. Asegurese de graficar el fiteo en el grafico anterior. [hint R: *lm* permite hacer regresiones lineales]
 - Considere la red de colaboraciones y la de internet nuevamente Encuentre cuantitativamente la asortatividad de la red utilizando ahora el estimador propuesto por Newman:

$$r = \frac{\sum_{ij}(A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij}(k_i \delta_{ij} - k_i k_j / 2m) k_i k_j}$$

Para ello tenga en cuenta lo desarrollado en las eqs [8.26 – 8.29] del libro de Newman. Como se corresponde este coeficiente con el estimado en el punto anterior? A qué se debe?

- Corra el script de cálculo (puntos i-iii) para las redes Y2H y AP-MS. Puede explicar lo que observa en cuanto a la asortatividad reportada?