

Data Mining en Ciencia y Tecnología

Pre-Tp 1: Grafos

Autor: Ricardo Amarilla

Comisión: 2

April 20, 2025

Abstract

El objetivo es el análisis de la topología de grafos del trabajo de Tagliazucchi y colaboradores (2013) que busca relacionar cambios en la modularidad de las redes construidas a partir de la señal de resonancia magnética funcional (fMRI) con los distintos estadios del sueño. Se pudo observar una semejanza con el modelo scale-free cuya distribución de grados sigue una ley de potencias permitiendo la existencia de hubs.

1 Introducción

El objetivo es el análisis de la topología de grafos en un área de investigación o un campo de estudio y el cálculo de parámetros básicos de un grafo y compararlos con modelos de redes random, small world y scale-free.

2 Resultados

Preprocesamiento de los datos — Objetivo 1: Reportar cuántos sujetos y cuántos estados de sueño se observan en el conjunto de datos.

Resolución: La cantidad de estados de sueño observados es de 4 y están conformados como ['N1', 'N2', 'N3', 'W']. La cantidad de sujetos únicos es de 18. Sujetos: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]

Manipulación de datos — Objetivo 1: Graficar la matriz de correlaciones entre regiones (es decir, la matriz de adyacencia ponderada) para el Sujeto 2 en la condición de vigilia ("Wake").

Resolución: Se utilizó la matriz de correlación obtenida a partir de los datos del sujeto 2. La visualización se realizó mediante un mapa de calor, que permite identificar patrones de conectividad entre regiones cerebrales. Ver la Figura 1a

Manipulación de datos — Objetivo 2: Transformar la matriz de adyacencia ponderada en una matriz de adyacencia binaria $A_{i,j}$ que represente una densidad de enlaces $\delta = 0,08$. ¿Cuál es el valor umbral de correlación entre pares de regiones que se debió utilizar?

Resolución: Para llevar a cabo la binarización de la matriz de adyacencia ponderada se utilizó un umbral del 0.77997. La visualización se realizó mediante un mapa de calor, que permite identificar patrones de conectividad entre regiones cerebrales de la matriz de adyacencia binarizada

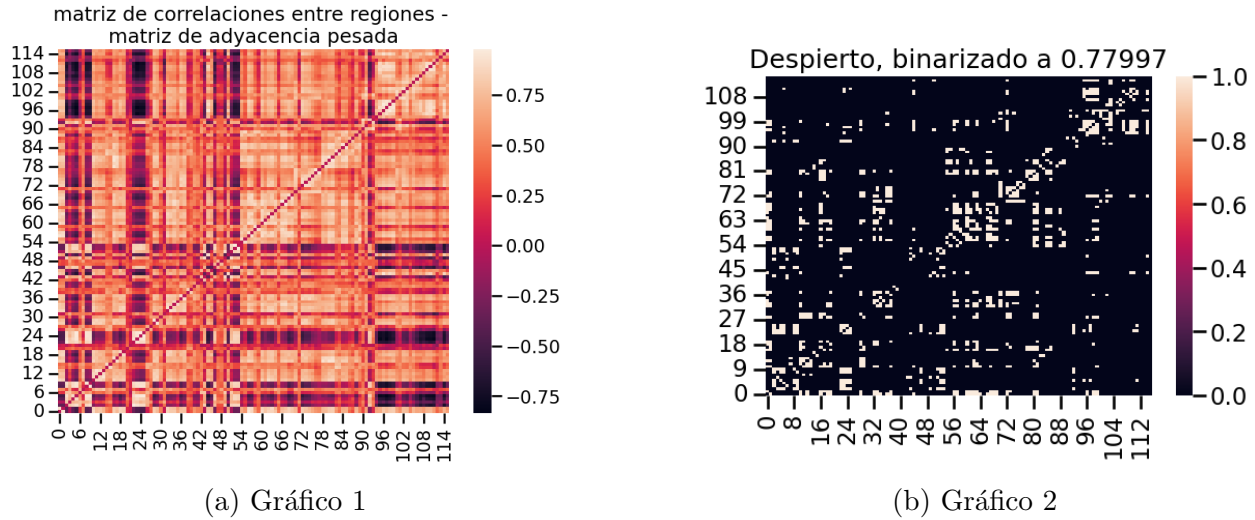


Figure 1: Distribucion de grados, Despierto

Manipulación de datos — Objetivo 3: Utilizando $A_{i,j}$, obtener el grafo resultante G .

Resolución: A partir de $A_{i,j}$ se obtuvo el grafo G . Ver la Figura 2a

Manipulación de datos — Objetivo 4: ¿Es G un grafo conectado? ¿Se puede calcular la distancia media entre pares de nodos $\langle d \rangle$ del grafo G ? En caso negativo, ¿qué medida equivalente podría utilizarse?

Resolución: El grafo G no está conectado por lo cual no es posible calcular la distancia media entre pares de nodos $\langle d \rangle$. La medida equivalente que sería interesante calcular es la Eficiencia Global que brinda la posibilidad de calcularse aún cuando algunos nodos no estén conectados.

Manipulación de datos — Objetivo 5: Calcular $\langle d \rangle$ para cada componente conexa de G . Calcular la eficiencia global eff del grafo G .

Resolución:

Se encontró que el grafo G posee 20 componentes de los cuales los relevantes son 3 el componente 1, 2 y 9, los restantes eran nodos aislados. A continuación los resultados obtenidos:

Componente	Nodos	Distancia Promedio (d)
1	92	3.117
2	5	1.600
9	2	1.000

Table 1: Distancia promedio por componente conexa

La Eficiencia global del grafo es del 0.264875597915335

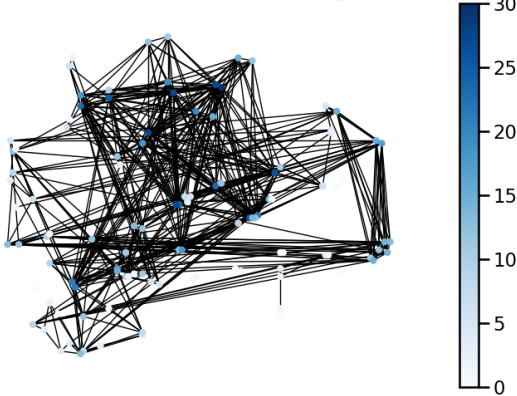
Manipulación de datos — Objetivo 6: Calcular el grado promedio $\langle k \rangle$, el nodo con grado máximo k_{\max} y el coeficiente de clustering promedio $C = \langle C_i \rangle$.

Resolución:

El grado promedio $\langle k \rangle$ es igual a 9.379310344827585, mientras que el nodo con grado máximo es el 0, y el k_{\max} es igual a 30. El coeficiente de clustering promedio $\langle C_i \rangle$ es igual a 0.5131233170907938

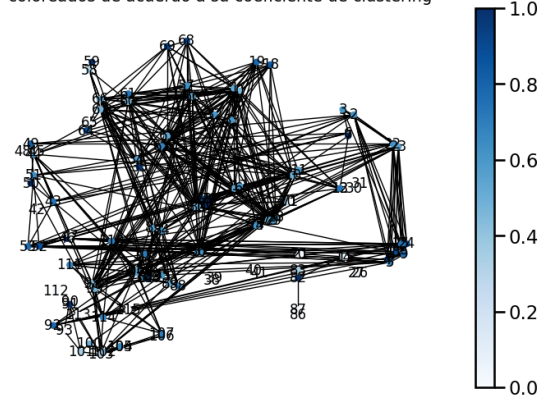
Manipulación de datos — Objetivo 7: Visualizar el grafo, ubicando los nodos en sus coordenadas cerebrales y coloreando cada nodo de acuerdo con su coeficiente de clustering C_i .

Con nodos coloreados de acuerdo a sus grados



(a) Gráfico 1

Con nodos en sus coordenadas cerebrales, coloreados de acuerdo a su coeficiente de clustering



(b) Gráfico 2

Figure 2: Grafo G con nodos en sus coordenadas cerebrales y coloreando cada nodo de acuerdo con su coeficiente de clustering y grados

Manipulación de datos — Objetivo 8: Graficar la distribución de grado del grafo (histograma normalizado), eligiendo un número adecuado de bins.

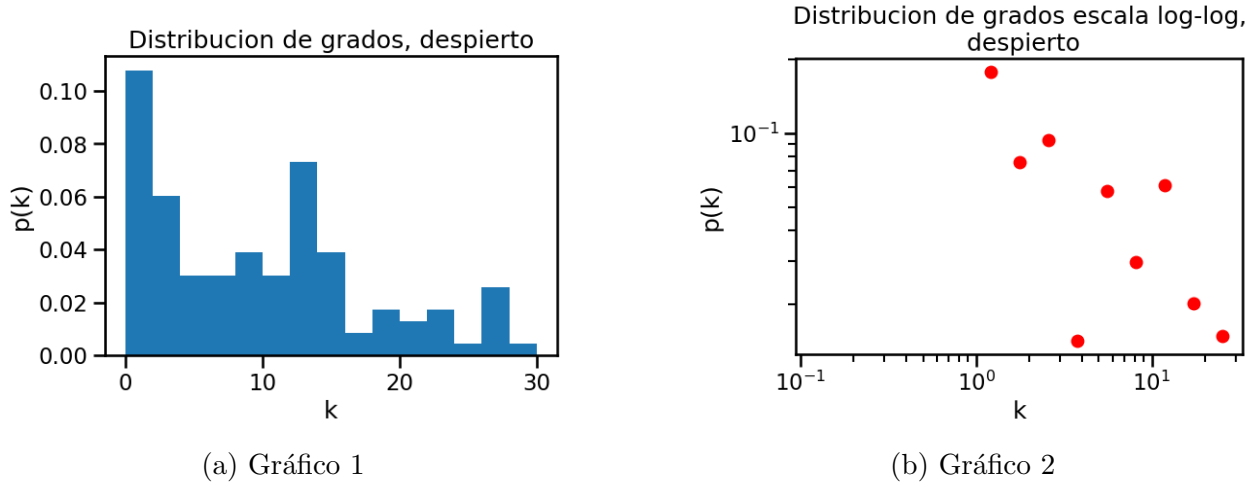


Figure 3: Distribucion de grados, Despierto

3 Conclusiones

Como se muestra en la Figura 3a y la Figura 3b, las distribución de grados se asemejan a una distribución que sigue una ley de potencias, los cuales generan grafos scale-free. Las conclusiones parecen ser las mismas cuando se analizan y comparan vigiliat extremas Wake y N3 para el mismo sujeto, las distribuciones se asemejan a una distribución de leyes de potencias utilizadas por los grafos scale-free con la consiguinete existencia de hubs lo cual es razonable siendo que las regiones del cerebro actuan siguiendo este patrón.

4 Referencias

- Barabási, A.-L. *Network Science*. Cambridge University Press, 2016.
- Tagliazucchi, E., Von Wegner, F., Morzelewski, A., Brodbeck, V., Borisov, S., Jahnke, K., & Laufs, H. *Large-scale brain functional modularity is reflected in slow electroencephalographic rhythms across the human non-rapid eye movement sleep cycle*. *NeuroImage*, **15**(1), 273–289, 2002.
- Romano, S., Varela, H., Bezchinsky, S., & Devouassoux, J. *Apuntes de Clase de Data Mining en Ciencia y Tecnología*. Facultad de Ciencias Exactas y Naturales, UBA, 2025.