
Pre TP2: Data Mining en Ciencia y Tecnología

José Saint Germain
joesg998@gmail.com

1 Introducción

El objetivo de este trabajo es familiarizarse con la generación de grafos que representen un conjunto de datos. A su vez, se buscará visualizar, manipular y comparar distintos tipos de grafos. Por último, se calculará parámetros básicos de un grafo y se los comparará con modelos de redes aleatorios (random), small world y libre de escala (scale-free).

2 Métodos

Nuestra fuente de trabajo es una base de datos de 18 sujetos en 4 estadios del sueño (despierto + 3 estadios del sueño). A su vez, para cada sujeto y estadio de sueño se utilizó una matriz de correlaciones de tamaño 116x116 con correlaciones entre las señales BOLD de 116 regiones cerebrales. En este artículo se trabajará con la matriz del sujeto número 12 en estado despierto (w).

Para procesar y visualizar los datos se utilizó el lenguaje de programación Python, del cual usamos las siguientes librerías: Numpy, para procesamiento de matrices; Pandas, para generación y procesamiento de datos tabulares; Matplotlib y seaborn, para visualización de datos y Networkx, para generación y manipulación de grafos.

3 Resultados y discusión

De manera exploratoria, se visualizó la matriz de correlaciones del sujeto 12 en estado despierto de manera pesada y binarizada.

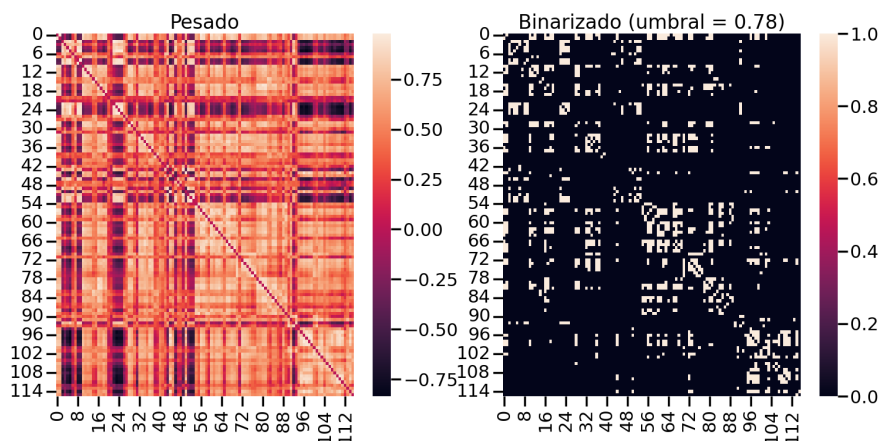


Figure 1: Matrices de correlación del cerebro del sujeto 2 despierto

18 La imagen de la izquierda muestra la matriz pesada, mientras que la de la derecha muestra la matriz
 19 binarizada con un umbral de 0.08. De esa manera, las correlaciones menores a ese umbral se muestra
 20 como negros y los mayores como blancos. De esta manera, se puede observar con claridad las
 21 correlaciones con valores bajos de toda la matriz.

22 Para continuar con el análisis, transformamos esta matriz en un grafo. El grafo resultante es una
 23 totalmente conectado, es decir que

Métrica	Valor
Distancia media	1.209295
Grado promedio (K)	11.500000
Nodo con grado máximo (kmax)	36.000000
Coefficiente de clustering promedio (C)	0.556270
Eficiencia	0.296510

Table 1: Métricas del grafo G

24 Subsecuentemente, observaremos el grafo simulando la forma de un cerebro, ubicando cada nodo en
 25 su ubicación correspondiente. Adicionalmente, agregamos un histograma de los grados de cada nodo.

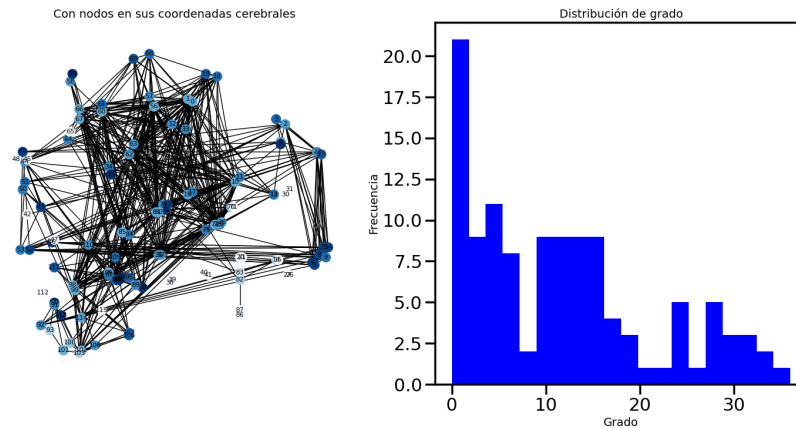


Figure 2: Visualización del grafo G

26 4 Conclusión

27 5 Bibliografía