

preTP1: Grafos

Sebastián Romano, Hernán Varela, Santiago Bezchinsky, Julián Devouassoux
Data Mining en Ciencia y Tecnología

31 de marzo de 2025

1. Introducción

El análisis de la topología de grafos (es decir, redes) es un área de investigación que atañe a diferentes campos de estudio. Para ejemplificar el uso de grafos nos enfocaremos en el los datos obtenidos en el 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 [1].

2. Objetivos

Familiarizarse con la generación de grafos que representen un conjunto de datos. Visualizar, manipular y comparar distintos grafos. Calcular parámetros básicos de un grafo, y compararlos con modelos de redes random, small world y scale-free.

3. Estructura de los datos:

Ejemplos de los procedimientos para comenzar el análisis pueden encontrarse en este enlace.

En la carpeta *DataSujetos* se encuentran los archivos separados por cada sujeto y estadio del sueño bajo la siguiente notación *[Estadio del sueño]_[Número de sujeto].csv*.

Para cada sujeto y estadio de sueño encontraremos una matriz de correlaciones de tamaño 116×116 con las correlaciones entre las señales BOLD de 116 regiones cerebrales.

Además se incluyen los nombres y coordenadas de las 116 regiones en un archivo aparte: *aal_extended_withCoords.csv*. Estas regiones están definidas a partir del atlas Automatic Anatomical Labeling (AAL) [2].

4. Preprocesamiento de los datos

- Cargar el dataset con los datos para cada sujeto y los nombres y coordenadas de las regiones cerebrales a las que se les registró la actividad.

- Reportar cuántos sujetos y cuántos estados de sueño se observan en el conjunto de datos.

5. Manipulación de datos

- Graficar la matriz de correlaciones entre regiones (es decir, la "matriz de adyacencia pesada") para el sujeto 2 de la condición despierto ("Wake")
- Transformar dicha matriz de adyacencia pesada a una matriz de adyacencia binaria $A_{i,j}$ que represente una densidad de enlaces $\delta=0.08$. ¿Cuál es el valor de umbral de correlación entre pares de regiones que tuvo que utilizar?
- Utilizando $A_{i,j}$, obtener el grafo resultante G
- ¿Es G un grafo conectado? ¿Se puede calcular la distancia media entre pares de nodos d del grafo G ? ¿Si no se puede, qué medida equivalente calcularías?
- Calcular d para cada componente conectado de G . Calcular la eficiencia global eff del grafo G .
- Calcular el grado promedio $\langle k \rangle$, el nodo con grado máximo k_{max} , el coeficiente de *clustering* promedio $C = \langle C_i \rangle$
- Visualizar el grafo, ubicando los nodos en sus coordenadas cerebrales y coloreando cada nodo de acuerdo a su coeficiente de *clustering* C_i
- Graficar la distribución de grado del grafo (el histograma normalizado, eligiendo un número de bins apropiado).

Referencias

- [1] Enzo Tagliazucchi, Frederic Von Wegner, Astrid Morzelewski, Verena Brodbeck, Sergey Borisov, Kolja Jahnke, and Helmut Laufs. Large-scale brain functional modularity is reflected in slow electroencephalographic rhythms across the human non-rapid eye movement sleep cycle. *Neuroimage*, 70:327–339, 2013.
- [2] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.