

# TP2: Redes en el Cerebro

Sebastian Romano, Juan E Kamienkowski, Hernán Varela, Álvaro López Malizia  
Data Mining en Ciencia y Tecnología

7 de noviembre de 2023

La aplicación del análisis de redes en general, y del análisis de grafos en particular, en registros de actividad cerebral en humanos es un campo en constante ebullición [Sporns, 2010, Barabási et al., 2016]. El primer paso del análisis es tomar una medida de similaridad entre la señal tomada de los electrodos o de las distintas regiones del cerebro. Estas medidas pueden estar enfocadas sobre una frecuencia particular o no, y suele tomarse en consideración variaciones de una sincronía de fase o a partir de las amplitudes [Sporns, 2010, Cohen, 2014].

Este trabajo está basado 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, detectados a partir de los ritmos de onda lenta en la señal del electroencefalograma (EEG) [Tagliazucchi et al., 2013]. Los estadios de sueño explorados no incluyen sueño REM (rapid eye movements) y fueron determinados manualmente por un experto según el criterio de la American Academy of Sleep Medicine (AASM) [Berry et al., 2012].

## 1. Objetivo

Este segundo Trabajo Práctico tiene tres objetivos:

- Explorar los cambios en la red en función de la profundidad del sueño.
- Explorar, en particular, los cambios asociados a la modularidad en función de la profundidad del sueño.
- Identificar los nodos en los cuales se producen estos cambios.

## 2. Estructura de los datos:

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*.

Además se incluyen los nombres de las 116 regiones en un archivo aparte: *aal\_extended.csv*. Estas regiones están definidas a partir del atlas Automatic Anatomical Labeling (AAL) [Tzourio-Mazoyer et al., 2002].

Ejemplos de los procedimientos para comenzar el análisis pueden encontrarse en

1. [https://colab.research.google.com/drive/1xU8p\\_YSeSxPAODgiyJwJAVuuDC-jrTwP#scrollTo=VG4joS9\\_OZCA%23offline%3Dtrue&sandboxMode=true](https://colab.research.google.com/drive/1xU8p_YSeSxPAODgiyJwJAVuuDC-jrTwP#scrollTo=VG4joS9_OZCA%23offline%3Dtrue&sandboxMode=true)
2. <https://colab.research.google.com/drive/1d0FDYATzQ0jm6tIt1inZuk6DMukEN5wS?usp=sharing%23offline%3Dtrue&sandboxMode=true#forceEdit=true&sandboxMode=true>

### 3. Tarea 1: Visualización

Para cada estadio del sueño (despierto - $W$ - y  $N1$ ,  $N2$ ,  $N3$ ) calcular el promedio entre sujetos de la matriz de correlaciones entre regiones cerebrales. Interpretaremos las matrices resultantes como la matriz de adyacencia pesada de un grafo donde los nodos son regiones cerebrales, para cada estadio del sueño. Visualizar dichas matrices. Vamos a definir un parámetro  $\delta$  que represente la densidad de enlaces, y en este TP lo haremos variar entre 0 y 0.15 (valores de densidad de enlaces fisiológicamente “realistas”). Con este parámetro  $\delta$  podremos generar grafos binarios con la densidad de enlaces deseada.

1. Primero comparar como varía el tamaño de la componente gigante (es decir, el porcentaje de nodos que participan en la componente gigante) en función de  $\delta$ , para todos los estadios. Interpretar las curvas obtenidas. ¿Qué nos indican los saltos?
2. Elegir uno de los saltos y visualizar los grafos correspondientes para valores de  $\delta$  justo antes y justo después del salto. Para esta, y todas las visualizaciones de grafos de este TP, representar siempre a los nodos en sus respectivas coordenadas cerebrales.
3. Graficar como varía el grado medio, el coeficiente de clustering medio, y la eficiencia global en función de  $\delta$ . Interpretar curvas.
4. Para un valor de  $\delta=0.12$  visualizar el grafo de cada estadio, donde el color del nodo  $i$  esté dado a su centralidad de autovector  $c_i^{eigen}$ , y su tamaño sea proporcional a  $\alpha c_i^{eigen}$  (donde  $\alpha$  es una constante que debemos elegir para ayudar la visualización). ¿Se observan cambios entre estadios?

### 4. Tarea 2: Comunidades y coeficiente de modularidad

Vamos a analizar como varía la modularidad entre los sujetos de cada estadio, en función de  $\delta$  (como siempre, en el rango  $\delta=0-0.15$ ). De aquí en adelante, siempre que se grafiquen curvas promedio, se deben graficar con una noción de incerteza (el error estándar de la media).

1. Para cada sujeto de cada estadio generar grafos variando  $\delta$ , determinar comunidades en dichos grafos a partir de algoritmo de Louvain [Blondel et al., 2008], y calcular la curva que describa el coeficiente de modularidad ( $Q$ ) en función de  $\delta$ . Promediando entre sujetos de cada estadio, obtener la curva correspondiente a cada estadio. Compararlas con las curvas de  $Q$  en función de  $\delta$  para “modelos nulos” equivalentes: usando grafos random de igual cantidad total de nodos e igual densidad de enlaces  $\delta$ . Para cada valor

de  $\delta$  incluir un test de significancia clásico para diferencias entre el grafo observado y el modelo nulo de grafo azar (marcar con un asterico las comparaciones significativas). Discutir los resultados.

2. Repetir el análisis número de comunidades ( $N_C$ ) encontradas en función de  $\delta$ .
3. Repetir el análisis de los dos puntos anteriores, pero ahora comparando específicamente el los datos del estado despierto  $W$ , con cada estadio  $N1$ ,  $N2$  y  $N3$ .
4. Usar las matrices de adyacencia pesadas promedio para cada estadio (definidas en la Tarea 1), fijar el  $\delta$  en un valor que les resulte interesante, encontrar la partición de módulos via Louvain y visualizar los grafos de cada estadio, coloreando cada nodo de acuerdo al módulo al que pertenece.

#### 4.1. Opcional 1:

Algoritmos de detección de comunidades Comparar el resultado del algoritmo de Louvain con otros, como Girvan-Newman. Discutir las las diferencias para el caso particular.

#### 4.2. Opcional 2:

Corrección por comparaciones múltiples. Plantee conceptualmente (es decir en palabras) las hipótesis estadísticas que se evaluaron con los tests de significancia de los puntos anteriores. ¿Qué es lo que se evaluó? Discuta específicamente si resulta necesario corregir por comparaciones múltiples para hacer afirmaciones sobre diferencias entre estadios, y cómo lo haría. De ser necesario, implementar una corrección que considere adecuada.

### 5. Tarea 3: Diferencias en las comunidades para los diferentes estadios

Para identificar diferencias significativas globales en las particiones de comunidades encontradas entre los distintos estadios del sueño ( $N1$ ,  $N2$  y  $N3$ ), con el estadio despierto ( $W$ ) se propone seguir el procedimiento propuesto por Alexander-Bloch y colaboradores [Alexander-Bloch et al., 2012]. Se quiere saber si las comunidades encontradas en un estadio de sueño  $NX$  y las encontradas en el cerebro despierto  $W$  son similares. Para cada estadio  $NX$  realizar el siguiente análisis para todo el rango de valores de  $\delta$ :

1. Se toman para todos los pares de sujetos en  $NX$  el Índice de Rand ajustado (adjusted-for-chance Rand index).
2. Se toman para todos los pares de sujetos en  $W$  el Índice de Rand ajustado (adjusted-for-chance Rand index).
3. Se promedian los valores obtenidos en los dos puntos anteriores, obteniéndose el Índice de Rand ajustado "*within-group*" observado ( $RI_o$ ). Se lo grafica junto al error estandar de la media.

4. Se aleatorizan las etiquetas de los pareos  $NX-W$  de forma que siempre esten todos los sujetos en cada grupo pero con las etiquetas cambiadas, y se repiten los pasos 1-3 de forma de obtener un Índice de Rand ajustado "*within-group*" permutado ( $RI_p$ ). Este paso se repite  $N_p$  veces generando una distribución empirica de valores de valores  $RI_p$ .
5. El p-valor se calcula como la cantidad de permutaciones que dieron  $RI_p > RI_o$  dividido  $N_p$ . Plantear un umbral de significancia y marcar con un asterico si la comparación fue significativa.

## 6. Opcional 3: Rol de nodos, y cambios en el rol de los nodos

Finalmente, se busca definir los roles que cumple cada nodo dentro de las comunidades, e identificar cuales son los nodos que, o bien están cambiando de comunidad o bien están cambiando de rol. Se deberán clasificar los nodos para cada estadio según el coeficiente de participación ( $P_i$ ) y el z-score del grado intra-comunidad ( $z_i$ ),

$$z_i = \frac{k_i - \langle k \rangle}{\sigma_k} \quad (1)$$

donde  $k_i$  es el grado intra-comunidad, y  $\langle k \rangle$  y  $\sigma_k$  son el promedio y el desvío estándar del grado intra-modular.

$$P_i = 1 - \sum_j^{N_M} \left( \frac{k_i^{U_j}}{k_i} \right)^2 \quad (2)$$

donde  $k_i^{U_j}$  es el número de aristas en el nodo  $i$  y la comunidad  $j$ ,  $N_M$  es el número de comunidades, y  $k_i$  es el grado total del nodo  $i$ .

Esto se realizará siguiendo los criterios propuestos en Tagliazucchi y colaboradores (2013) [Tagliazucchi et al., 2013]:

- Hubs:  $P_i > P_C$  &  $z_i > z_C$
- Provincial Hubs:  $P_i < P_C$  &  $z_i > z_C$
- Provincial Nodes:  $P_i < P_C$  &  $z_i < z_C$
- Connector Nodes:  $P_i > P_C$  &  $z_i < z_C$

en principio, para los umbrales propuestos ( $P_C = 0,05$  y  $z_C = 1$ ) en el mismo trabajo. Si se anima, puede definir otros umbrales a partir del análisis de los histogramas de  $P_i$  y  $z_i$ .

1. Graficar el número de nodos por cada clase en función de  $\delta$ , y comparar estadísticamente cada estadio  $N1$ ,  $N2$  y  $N3$ , con el estadio despierto  $W$ .
2. Nuevamente usando las matrices de adyacencia pesadas promedio para cada estadio (definidas en la Tarea 1), elegir un valor de  $\delta$  que resulte particularmente interesante a partir de los análisis anteriores, y visualizar los Hubs y los Provincial Hubs en el grafo coloreando el nodo y sus enlaces.

## 7. Formato

Repetir el formato anterior incorporando los comentarios de la devolución del TP1.

## 8. Nota final

El TP se realizará en grupos de tres a cinco personas. El TP consiste de una serie de tareas, que pueden consistir en un análisis o contestar una pregunta. Algunas de estas preguntas o tareas están indicadas como optativas. Realizar estas tareas suma puntos pero no son obligatorias. Se puede usar cualquier herramienta de análisis o combinación de herramientas, debiendo indicarla en el informe. El lenguaje en el que se desarrolle el TP no es excluyente.

Tabla de puntos: Cantidad máxima de puntos que se pueden obtener por ...

- ... la tarea obligatoria 1: 1.5
- ... la tarea obligatoria 2: 2.5
- ... la tarea obligatoria 3: 2.5
- ... la tarea opcional 1: 1.0
- ... la tarea opcional 2: 1.0
- ... la tarea opcional 3: 1.5

Puntaje máximo posible: diez

## Referencias

- [Alexander-Bloch et al., 2012] Alexander-Bloch, A., Lambiotte, R., Roberts, B., Giedd, J., Gogtay, N., and Bullmore, E. (2012). The discovery of population differences in network community structure: new methods and applications to brain functional networks in schizophrenia. *Neuroimage*, 59(4):3889–3900.
- [Barabási et al., 2016] Barabási, A.-L. et al. (2016). *Network science*. Cambridge university press.
- [Berry et al., 2012] Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., Vaughn, B. V., et al. (2012). The aasm manual for the scoring of sleep and associated events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 176.
- [Blondel et al., 2008] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.
- [Cohen, 2014] Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT press.
- [Sporns, 2010] Sporns, O. (2010). *Networks of the Brain*. MIT press.
- [Tagliazucchi et al., 2013] Tagliazucchi, E., Von Wegner, F., Morzelewski, A., Brodbeck, V., Borisov, S., Jahnke, K., and Laufs, H. (2013). Large-scale brain functional modularity is

reflected in slow electroencephalographic rhythms across the human non-rapid eye movement sleep cycle. *Neuroimage*, 70:327–339.

[Tzourio-Mazoyer et al., 2002] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289.