

Data Mining en Ciencia y Tecnología: Trabajo Práctico 2

Marcelo Soria - Juan Kamienkowski

Noviembre 2018

Indicaciones para la realización del TP:

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 indicarlo en el informe. Los ejemplos de esta guía están en R, pero eso no es excluyente.

La fecha de entrega para la segunda parte es el **16 de Diciembre** hasta las 24 hs, para ambos turnos.

Tabla de puntos:

Cantidad máxima de puntos que se pueden obtener por la tarea obligatoria 1: 1.0

Cantidad máxima de puntos que se pueden obtener por la tarea obligatoria 2: 1.0

Cantidad máxima de puntos que se pueden obtener por la tarea obligatoria 3: 2.0

Cantidad máxima de puntos que se pueden obtener por la tarea obligatoria 4: 2.5

Cantidad máxima de puntos que se pueden obtener por la tarea obligatoria 5: 2.5

Cantidad máxima de puntos que se pueden obtener por la tarea opcional 1: 0.5

Cantidad máxima de puntos que se pueden obtener por la tarea opcional 2: 1.0

Cantidad máxima de puntos que se pueden obtener por la tarea opcional 3: 0.5

Cantidad máxima de puntos que se pueden obtener por la tarea opcional 4: 0.5

Cantidad máxima de puntos que se pueden obtener por la tarea opcional 5: 0.5

Puntaje máximo posible: diez

Introducción:

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). 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, 2007).

Objetivos:

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.

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]_suj[Número de sujeto].csv*.

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

Tarea 1: Visualización

A partir de los datos individuales, visualizar la estructura de las redes pesadas a partir de los datos promedio para cada estadio del sueño (despierto -**W**- y **N1**, **N2**, **N3**).

Transformar el grafo pesado de los datos promedio en uno no pesado manteniendo constante la densidad de aristas (δ). Sobre los datos promedio, extraer medidas de centralidad, grado, camino mínimo, y coeficiente de clustering en función de la densidad de aristas en función de δ .

Visualizar el grafo no pesado para distintos valores de δ que resulten interesantes.

Tarea 2: Comunidades y coeficiente de modularidad

Para los distintos valores de la densidad de aristas (δ), determinar comunidades a partir de algoritmo de Louvain (Blondel et al, 2008) y estimar el coeficiente de modularidad (**Q**). Graficar **Q** y el número de comunidades (**N_c**) en función de δ , y comparar los resultados con el comportamiento de una red random que preserve la distribución de grados de los nodos.

Opcional 1 (Tarea 2): Algoritmos de detección de comunidades Comparar el resultado del algoritmo de Louvain con otros, como Girvan-Newman. Discutir las diferencias para el caso particular.

Tarea 3: Estadística

Repetir las curvas de modularidad (**Q**) y número de comunidades (**N_c**) para cada sujeto y estadio del sueño, en función de δ , y comparar cada estadio **N1**, **N2** y **N3**, con el estadio despierto **W**. Para cada valor de δ incluir una noción de significancia entre estadios del sueño.

Opcional 2 (Tarea 3): Corrección por comparaciones múltiples En cada caso discuta si resulta necesario corregir por comparaciones múltiples y cómo lo haría. De ser necesario, implementar una corrección adecuada.

Opcional 3 (Tarea 3): Corrección por comparaciones múltiples Repetir el procedimiento sobre las medidas de centralidad, grado, camino mínimo, y coeficiente de clustering ¿Cómo interpretaría estos resultados a la luz de los obtenidos para la modularidad?

Tarea 4: Diferencias en la membresía para los diferentes estadios

Para identificar diferencias significativas globales en la membresía de los nodos 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). Para cada par **NX** y **W** se calcula la similitud entre grupos a partir de Índice de Rand ajustado (*adjusted-for-chance Rand index*), y el valor promedio se compara con Np permutaciones de las etiquetas de comunidades para cada estadio. El p-valor se obtiene dividiendo el número de veces en que una permutación supera el valor promedio para los datos no permutados por el número total de permutaciones (Np).

Tarea 5: 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 σ_k son el promedio y el desvío estándar del grado intra-modular.

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

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):

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 y z_C) en el mismo trabajo. Graficar el número de nodos por cada clase en función de δ , y comparar estadísticamente cada estadio **N1**, **N2** y **N3**, con el estadio despierto **W**.

Para un valor de δ que resulte particularmente interesante a partir de los análisis anteriores, visualizar el grafo coloreando la membresía de cada nodo y su rol. Identificar en este grafo cuáles son los nodos que cambian entre los distintos estadios del sueño.

Opcional 4 (Tarea 5): Criterios para definir los roles Revisar los criterios definidos por Tagliazucchi y colaboradores (2013) a partir de los histogramas del *coeficiente de participación* (P_i) y del *z-score del grado intra-comunidad* (z_i).

Opcional 5 (Tarea 5): Visualizaciones Usar el archivo *aal_extended.csv* para generar visualizaciones coloreando por regiones anatómicas y hemisferios. ¿Qué puede concluir a partir de las diferencias entre agrupamientos funcionales y anatómicos para los diferentes estadios del sueño?

Referencias:

- AASM, 2007. The AASM manual for the scoring of sleep and associated events-rules. Terminology and technical specifications. American Academy of Sleep Medicine, Chicago.
- Alexander-Bloch, A., Lambiotte, R., Roberts, B., Giedd, J., Gogtay, N., Bullmore, E., 2012. The discovery of population differences in network community structure: new methods and applications to brain functional networks in schizophrenia. *Neuroimage* 59, 3889–3900.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. *J. Stat. Mech.* P10008.
- Tagliazucchi, E., Von Wegner, F., Morzelewski, A., Brodbeck, V., Borisov, S., Jahnke, K., & 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, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & 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.