
Grafos en neurociencias (pre TP2)

Víctor A. Bettachini

Datamining en ciencia y tecnología 2023

Especialización en Explotación de Datos y Descubrimiento del Conocimiento

bettachini@gmail.com

1. Materiales y métodos

Datos Se hace uso de datos producto de la medición de la señal de resonancia magnética funcional (fMRI). Definidos 116 volúmenes de interés del cerebro en términos de su activación [1], se publicaron coeficientes de correlación lineal entre sus medias en distintos segmentos temporales [2]. Con estos datos se generó una matriz de correlación, y a partir de la misma los grafos analizados en este trabajo.

Recurso informático Un cuaderno (notebook) Jupyter provisto por los docentes en el sitio web denominado “Campus” [3] es la plantilla donde se escribió código en lenguaje Python. Este explotó funciones de las biblioteca NetworkX [4].

1.1. Preprocesamiento de los datos

Carga del conjunto de datos Los archivos provistos corresponden a los estadios de sueño N1, N2, N3 y despierto (W) para 18 sujetos. Estos estuvieron acompañados de una tabla que describe la denominación y ubicación espacial las regiones en que se parcializó el cerebro.

2. Resultados

2.1. Manipulación de datos

Matriz de correlación La matriz de adyacencia pesada que muestra la figura 1a corresponde a la condición de despierto para el sujeto número 2.

Una medida que caracteriza un grafo es la densidad, δ , definida como la razón entre los enlaces presentes sobre todos los posibles en el grafo. Para convertir la matriz en una de adyacencia binaria con una densidad de enlaces $\delta = 0,8$ se discriminaron sus pesos con un umbral 0,77997 y así se obtuvo la matriz que muestra la figura 1b.

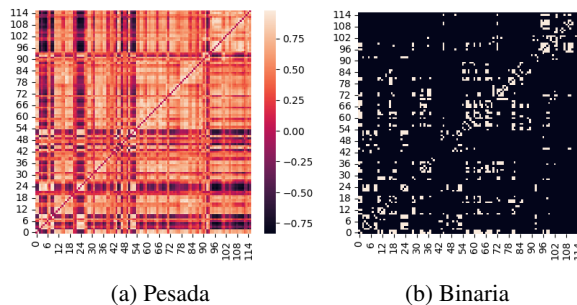


Figura 1: Matrices de correlación de las medias de señales para el sujeto número 2 despierto.

Grafo de la matriz de adyacencia binaria No es totalmente conectado. Hay tres componentes conectadas de 92, 5 y 2 nodos en tanto que los 17 restantes están aislados. La distancia mínima entre nodos contabiliza cuantos intermedios deben atravesarse para ir de un nodo a otro. Esta medida para los aislados no tiene sentido, por lo en este grafo no puede obtenerse una distancia media d a lo fines de tener una media de que tan “conectado” se presenta el grafo. Una medida numérica alternativa la provee la densidad. Para el conjunto, no conectado, este valor parece bajo, $\approx 8\%$. Y en el componente mayoritario, el de 92 nodos, este valor apenas se incrementa hasta un $\approx 12\%$.

Eficiencia de conectividad La suma de la inversa de es una medida de la eficiencia de conectividad global del grafo, $\text{eff} = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d(i,j)}$ [5]. No logré realizar un cálculo manual en línea con valores publicados para de esta medida, pero la biblioteca NetworkX provee una función que la calculó en $\approx 0,245$.

Distribución de grado El número de enlaces por nodo, o grado k , se distribuye en forma dispar. De un total de 534 enlaces el grado mayor resultó $k_{\text{máx}} = 30$, y un relativamente alto promedio $\langle k \rangle \approx 9,207$ aunque no hay que olvidar que no participan aquí los nodos aislados. El histograma de k que muestra la figura 2a que este $\langle k \rangle$ es representativo de la distribución.

En una medida similar a la densidad que cuenta la proporción de enlaces sobre los posibles puede hacerse algo similar calculando la proporción de cuantos de los posibles enlaces entre primeros vecinos efectivamente se realizan. Esto se denomina coeficiente de agrupamiento (*clustering*) por nodo C_i , cuyo promedio para este grafo es $\langle C_i \rangle \approx 0,527$. Coloreando cada nodo según su C_i y ubicándolo por las coordenadas (y,z) en el cerebro se obtiene la figura 2b.

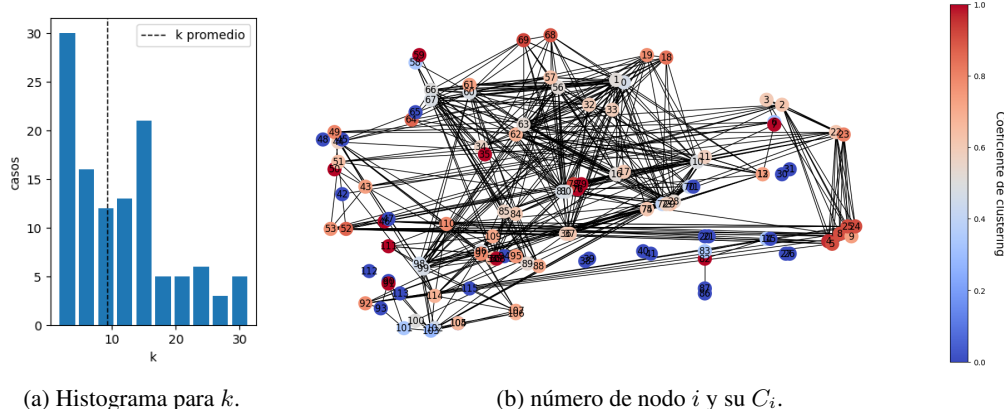


Figura 2: Distribución de grado y coeficiente de agrupamiento.

Grafos prototípicos Tres algoritmos generaron grafos de conexiones entre el mismo número de nodos, $n = 116$, que de la matriz de adyacencia binaria. Se buscó que estos presente coeficientes de caracterización similares al del grafo de esta.

El algoritmo de Erdős–Rényi enlaza un par de nodos si una probabilidad al azar supera un umbral [6, sección 3.2]. Estableciendo en función del $\langle k \rangle$ de la red original tal probabilidad como $p = \frac{\langle k \rangle}{n-1}$ se obtuvo una densidad prácticamente idéntica $\delta = 0,0808$ a la de la red original, $0,0801$, con una suma de enlaces, 539, también muy cercana, 534. El coeficiente de agrupamiento $\langle C_i \rangle = 0,7986$ resulta bastante disímil con el $0,5271$ de la red original. Esta red azarosa, a diferencia de la original, mostró estar totalmente conectada, como se aprecia en la figura 3a. En contrapartida, las redes reales suelen estar particionadas en múltiples componentes conectadas [6, sección 3.7].

Si una red fuera perfectamente regular y todos los nodos tuvieran idéntica k , la distancia mínima entre dos nodos d , es decir cuantos intermedios debe atravesarse para enlazar uno con otro, seguiría una dependencia polinomial con n . En las redes reales d presenta una dependencia mucho menor con n , con $\log(n)$ de hecho. Este fenómeno recibe el nombre de *mundo pequeño*, o *small-world*, por la sorpresiva poca distancia entre dos cualesquiera nodos de la red. Asimismo, en las reales, el coeficiente de agrupamiento C_i suele ser más alto que en una red azarosa [6, sección 3.9]. En

respuesta a estas observaciones, el **algoritmo de Watts-Strogatz** extiende el modelo de un k regular con una probabilidad de que un enlace cambie a nodo, lo que genera redes intermedias entre una regular y una azarosa. El número de enlaces total más cercano al de la red original, 580, se obtuvo con $k = \text{entero}(\langle k \rangle) + 1$. Puesto que la $\delta = 0,870$ es insensible al parámetro de reconexión se lo ajustó buscando un $\langle C_i \rangle$ similar al de la red original. Este fluctúa por efectos azarosos pero se logró que este oscilara en torno al $\approx 0,527$ con una probabilidad de reconexión de 0,815. Pero la red generada se presenta totalmente conectada y con un aspecto muy diferente a la original, como muestra la figura 3b. Se pudo generar una red con aspecto más similar incrementando la reconexión y así haciéndola más azarosa y similar a la de las mediciones, pero el resultante $\langle C_i \rangle$ se aleja del de esta.

Hasta aquí había esquivado el tratar la distribución de los grados k según el modelo prototípico de red. El proceso binomial de una red azarosa para un $n \gg \langle k \rangle$ presenta una distribución de Poisson [6, sección 3.4]. Se comentó ya el fenómeno de *mundo pequeño*, pero no el los menores d son fruto de la existencia de nodos con alto k llamados *hubs*. La distribución de probabilidad de k se extiende mostrando frecuentemente una dependencia con una ley de potencias $p_k \sim k^{-\gamma}$, dando origen al designación de redes *sin escala* o *scale-free* [6, sección 4.2]. El **algoritmo de Barabási-Albert** genera tales redes a partir de un parámetro m que determina cuantos enlaces se agregan en cada paso. Se ajustó $m = \text{entero}(\frac{\langle k \rangle}{2}) + 1$ para dar cuenta de no contar dos veces el número medio de enlaces de la red original y el uno adicional se agregó para que la red resultante tuviera el número de enlaces más similar a esta (555). Lo generado presentado en la figura 3c también muestra una red conectada, pero con una similitud adicional con la original al mostrar claros *hubs*.

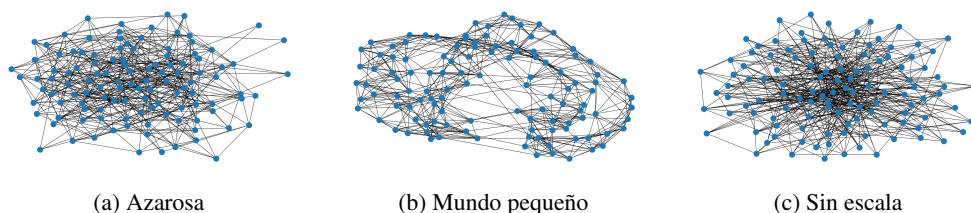


Figura 3: Grafos generados respetando el número de nodos y aproximadamente el de enlaces de la red original con algoritmos de (a) Erdős-Rényi, (b) Watts-Strogatz y (c) Barabási-Albert.

Coefficientes de grafos prototípicos Las redes prototípicas comentadas en la sección anterior son una realización al azar. Para dar respuesta a que tan similares son los coeficientes que les caracterizan a los de la red producto de las mediciones se hicieron 1000 corridas de estos algoritmos a los fines de realizar histogramas de del número medio de enlaces, $\langle k \rangle$, el máximo número de estos, $\text{máx}k$, el coeficiente de agrupamiento medio, $\langle C_i \rangle$, y la eficiencia de conectividad, eff .

En el caso de las redes azarosas generadas con el algoritmo de Erdős-Rényi, la figura 4 muestra que el $\langle k \rangle$ y los $\text{máx}k$ los exceden ligeramente. Pero los $\langle C_i \rangle$ están muy por debajo y la eff por arriba de los de la red original indicados con líneas punteadas verticales. Lo primero no es extraño pues de hecho las redes al azar presentan sistemáticamente un $\langle C_i \rangle$ menor que las reales [6, sección 3.9]. Para lo segundo espera un resultado opuesto, pues en las redes reales d son menores que en la red azarosa esperaba que la suma de sus inversas fuera menor que en la red producto de las mediciones. Pienso que estoy cometiendo un error aquí.

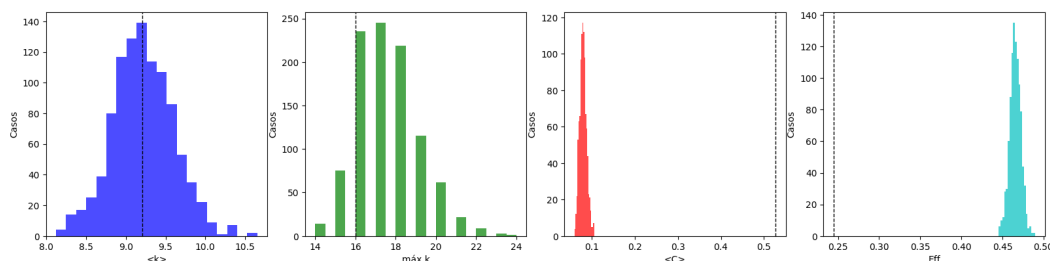


Figura 4: Magnitudes de las redes azarosas.

En el caso de las redes generadas con el algoritmo de Watts-Strogatz, la figura 5 muestran un $\langle k \rangle$ superior al de la red producto de las mediciones pero unos $\text{máx } k$ comparables. Puesto que se ajustó para que así fuera la concordancia con el coeficiente de agrupamiento medio es buena. Nuevamente la eficiencia global arroja un valor mucho mayor que el de la red original que no puedo explicar. Si no fuera por este desajuste la red del tipo *mundo pequeño* sería una buena candidata para representar el fenómeno que dio origen a la red original.

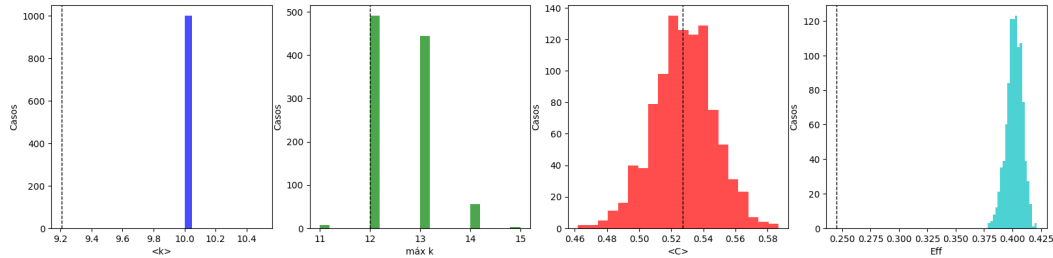


Figura 5: Magnitudes de las redes *mundo pequeño*.

En el caso de las redes generadas con el algoritmo de Barabási-Albert, la figura 6 muestran un $\langle k \rangle$ superior al de la red producto de las mediciones y unos $\text{máx } k$ un tanto mayores pero comparables. En contrapartida el coeficiente de agrupamiento medio es mucho menor que el de la red original. Y nuevamente no puedo explicar la eficiencia global mayor que la de la red original.

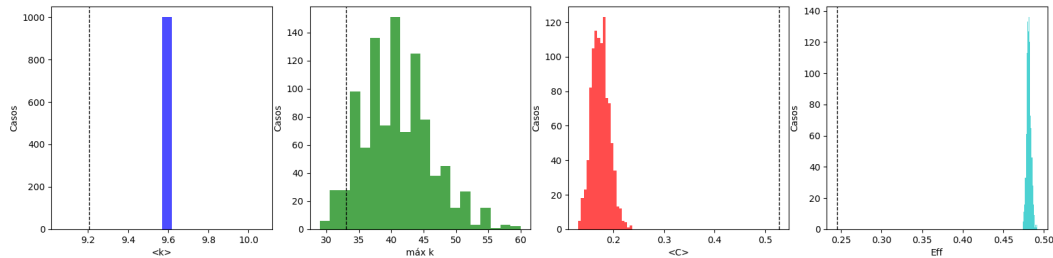


Figura 6: Magnitudes de las redes *sin escala*.

Referencias

- [1] N. Tzourio-Mazoyer et al. «Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain». En: *NeuroImage* 15.1 (ene. de 2002), págs. 273-289. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1053811901909784>.
- [2] Enzo Tagliazucchi et al. «Large-scale brain functional modularity is reflected in slow electroencephalographic rhythms across the human non-rapid eye movement sleep cycle». En: *NeuroImage* 70 (abr. de 2013), págs. 327-339. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1053811913000128>.
- [3] Juan A. Kamienkowski. *Curso: Data Mining en Ciencia y Tecnología*. 2023. URL: <https://datamining.dc.uba.ar/campus/course/view.php?id=37>.
- [4] Aric A Hagberg, Daniel A Schult y Pieter J Swart. «Exploring Network Structure, Dynamics, and Function using NetworkX». En: *Proceedings of the 7th Python in Science Conference (SciPy 2008)*. SciPy 2008. Pasadena, CA USA, ago. de 2008, págs. 11-15. URL: https://conference.scipy.org/proceedings/SciPy2008/paper_2/.
- [5] Bryan Ek, Caitlin VerSchneider y Darren A. Narayan. «Global efficiency of graphs». En: *AKCE International Journal of Graphs and Combinatorics* 12.1 (1 de jul. de 2015), págs. 1-13. URL: <https://www.tandfonline.com/doi/full/10.1016/j.akcej.2015.06.001>.
- [6] Albert-László Barabási. *Network Science*. Cambridge University Press, jul. de 2016. 456 págs. URL: <http://networksciencebook.com/>.