

Análisis de distancias

A continuación se muestran las gráficas de los P-values calculados para los dos coeficientes (Clustering Coefficient y Average Minimum Path Length) fijando el umbral de binarización del grafo para que mantenga el 15% de las aristas más positivas, que es el caso óptimo encontrado por Roser y verificado después por nosotros.

Todos las gráficas muestran P-value (eje Y) en función de un parámetro que ajusta la distancia máxima o mínima permitida, según el caso (eje X). Este estudio se ha realizado en 2 escenarios diferentes:

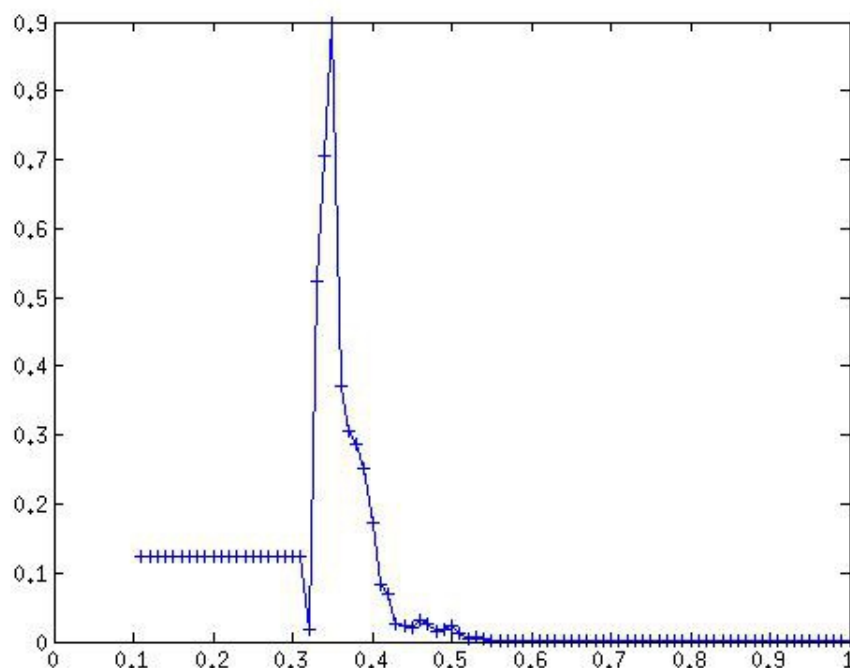
1. Barrido que elimina todas las aristas que representan la unión de dos regiones **cuya distancia es superior al parámetro**. Barrido de distancia 0 (elimina todas las aristas) hasta 1 (admite todas las aristas). Barrido en saltos de 0.01.
2. Barrido que elimina todas las aristas que representan la unión de dos regiones **cuya distancia es inferior al parámetro**. Barrido de distancia 0 (admite todas las aristas) hasta 1 (elimina todas las aristas). Barrido en saltos de 0.01.

IMPORTANTE: Por eliminar una arista se entiende que, al no cumplir la condición de distancia, la arista pasa automáticamente a tener un valor que nunca la hará seleccionable para estar en el 15% de aristas con valor más alto, pero sigue contando a la hora de calcular qué número de aristas respecto al total es ese 15%. Así, después de binarizarlo, el grafo resultante tiene el mismo número de aristas que el original.

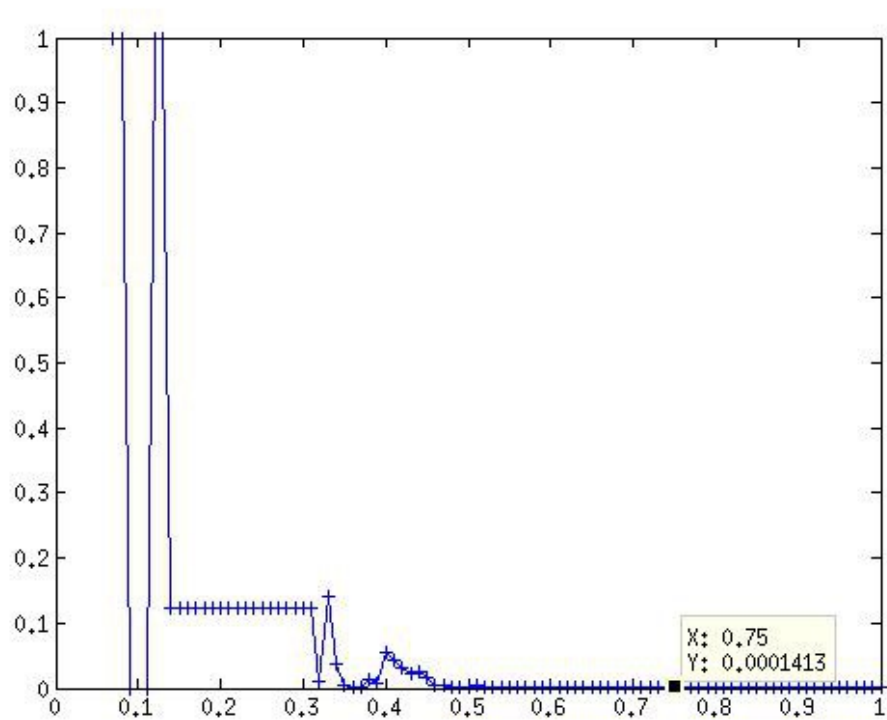
Creo que mantener el mismo número de aristas hace que comparar los descriptores sea más significativo. No estoy muy seguro de esto, pero comparar grafos con distinto numero de aristas no suena muy natural porque tanto el clustering coefficient como al average minimum path length dependen del numero de aristas que tenga el grafo, y lo que queremos ver es como se parecen estos grafos al original. ¿Cómo lo ves?.

Escenario 1 (eliminación de distancias mayores al parámetro):

Clustering Coefficient

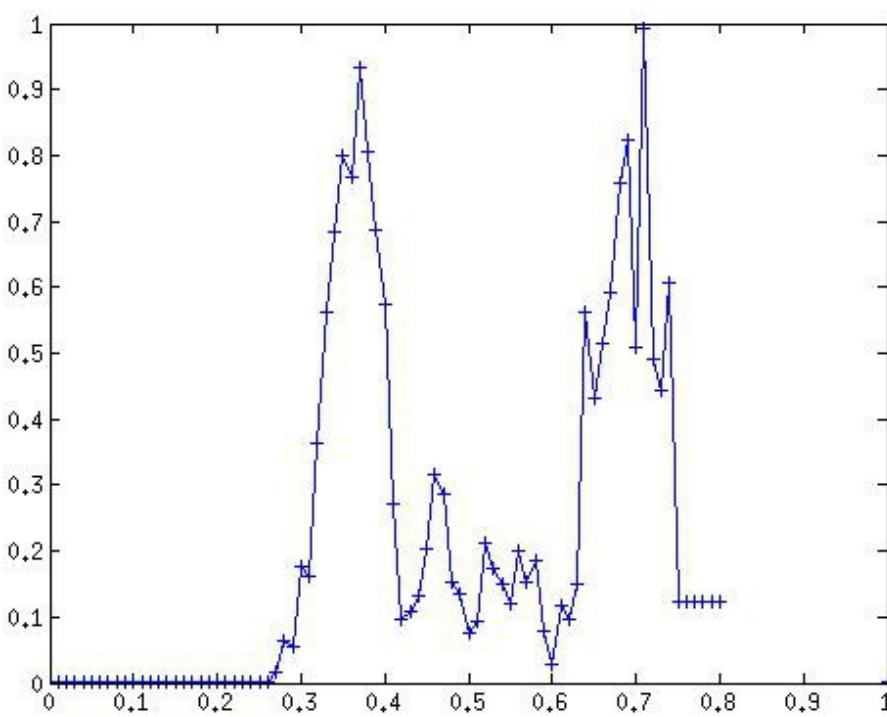


Average Minimum Path Length

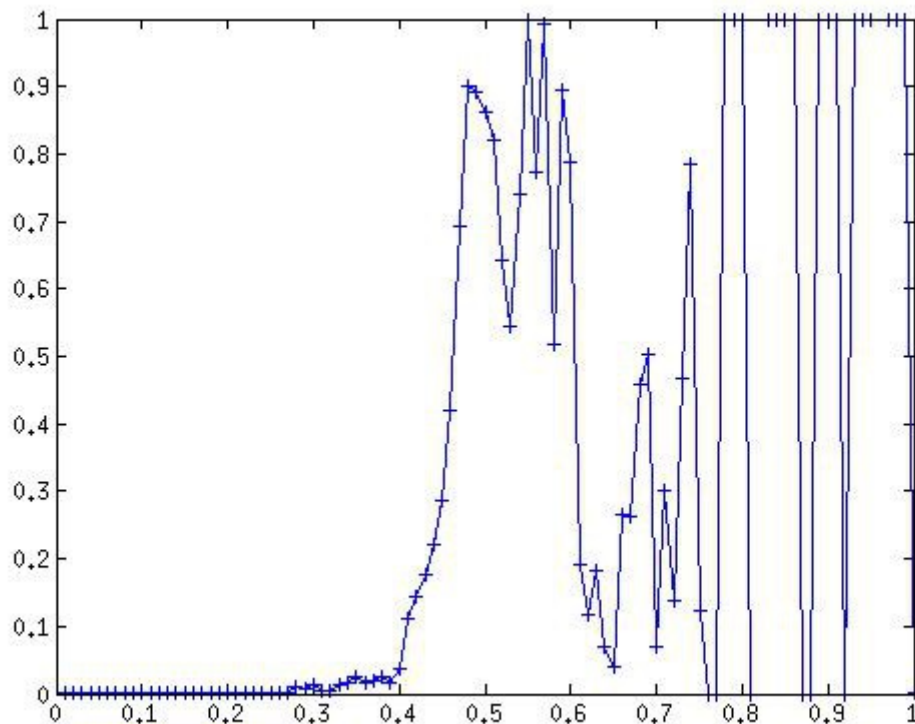


Escenario 2 (eliminación de distancias menores al parámetro):

Clustering Coefficient



Average Minimum Path Length



Resultados

Eliminación de distancias mayores al parámetro

CC: Eliminando distancias mayores al parámetro a partir de 0.55 y hasta 1.0, el valor P se mantiene por debajo de 0.001 (mínimo de $1.068e-07$ y máximo de 0.000663). Esto implica que las aristas correspondientes a distancias menores a 0.55 son las que aportan la información.

AMPL: Eliminando distancias mayores al parámetro a partir de 0.54 y hasta 1.0, el valor P se mantiene por debajo de 0.001 (mínimo de $6.162e-05$ y máximo de 0.0005801). Esto implica que las aristas correspondientes a distancias menores a 0.54 son las que aportan la información.

Para distancias mayores a 0.54 se eliminan 1932 aristas de las 4005 totales.

Eliminación de distancias inferiores al parámetro

CC: Eliminando distancias mayores al parámetro a partir de 0 y hasta 0.26, el valor P se mantiene por debajo de 0.001 (mínimo de $3.803e-07$ y máximo de 0.000875). Esto implica que las aristas correspondientes a distancias mayores a 0.26 son las que aportan la información.

AMPL: Eliminando distancias mayores al parámetro a partir de 0 y hasta 0.24, el valor P se mantiene por debajo de 0.001 (mínimo de $9.658e-05$ y máximo de 0.0004542). Esto implica que las aristas correspondientes a distancias mayores a 0.24 son las que aportan la información.

Para distancias menores a 0.26 se eliminan 368 aristas de las 4005 totales.

Comentarios y conclusiones

Parece que **las aristas que aportan la mayor parte de la información** son las que unen regiones del cerebro cuyas **distancias están comprendidas entre 0.26 y 0.54**, que es cuando más se parecen los grafos al grafo original.

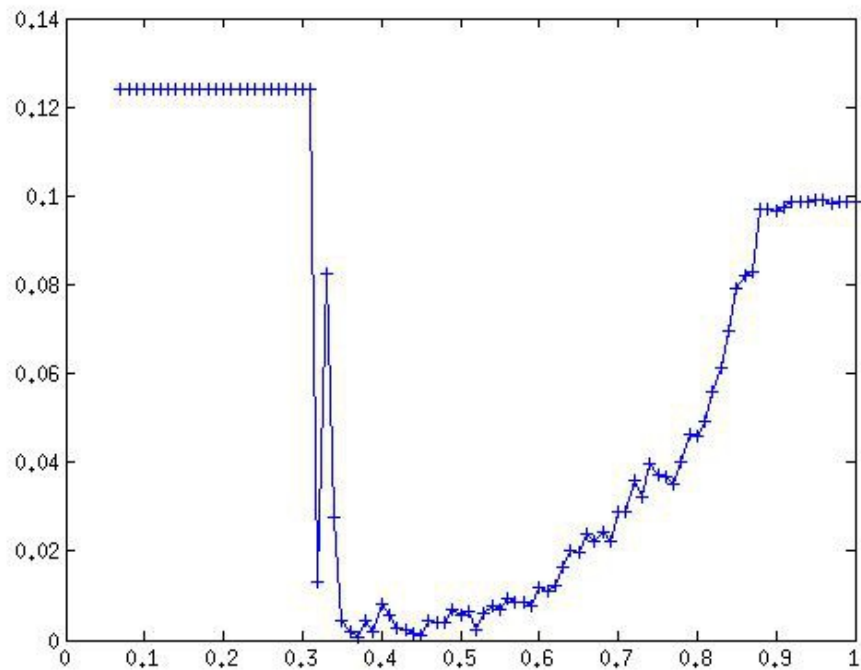
Esto me parece interesante, sobretodo porque por encima de 0.54 hay 1932 aristas, que es casi la mitad del total. Sumando las que hay por debajo de 0.26 con las que hay por encima de 0.54 tenemos un total de 2300 aristas. El total de aristas es de 4005. **¿Implica esto que un 57% de las aristas se pueden eliminar fijándose sólo en la distancia sin que esto afecte a la correlación de los descriptores con la edad?**

Anexo

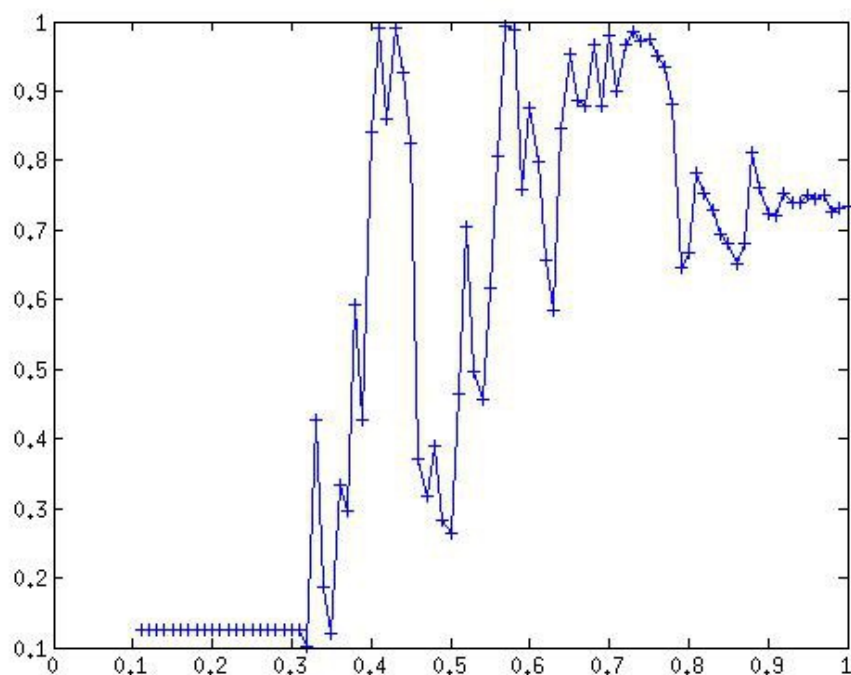
Comportamiento de los otros 2 descriptores (Global Efficiency y Small-World) en cada uno de los escenarios. Siempre mal comportamiento, irrelevantes como en el grafo original.

Escenario 1 (eliminación de distancias mayores al parámetro):

Global Efficiency

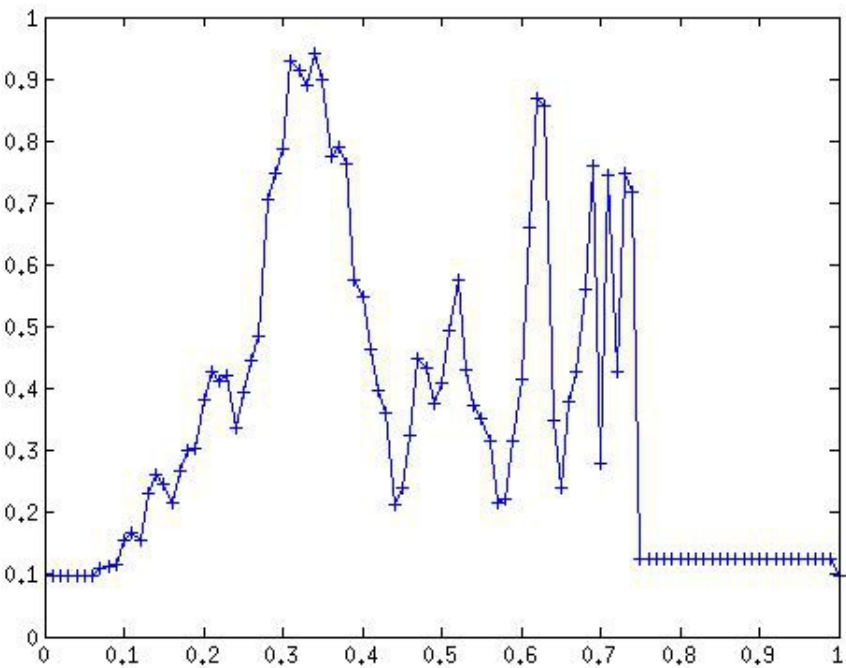


Small-World



Escenario 2 (eliminación de distancias menores al parámetro):

Global Efficiency



Small-World

