



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y
DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

DISTANCIA SIN PROPAGACIÓN HACIA ATRÁS ENTRE GRAFOS CON CONO

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRA EN CIENCIAS MATEMÁTICAS

PRESENTA:
LIUBOVE ORLOV SAVKO

DIRECTOR DE TESIS
DR. PABLO SUÁREZ SERRATO
INSTITUTO DE MATEMÁTICAS, CU

Ciudad Universitaria, Cd. Mx. Noviembre 2020



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Índice general

Agradecimientos	ii
Introducción	iii
1 Preliminares	1
1.1 Topología algebraica	1
1.2 Teoría de grafos	5
1.3 Ciclos sin retrocesos	7
1.3.1 Espectro de la matriz sin retrocesos	11
1.4 Estadística	13
1.4.1 Estimación de densidad con núcleo	16
1.5 Transporte Óptimo	16
1.5.1 Relajación de Kantorovich	20
1.5.2 Distancia de Gromov-Wasserstein	23
2 Inferencia de redes de fMRI	25
2.1 Descripción de los datos	25
2.2 Inferencia de conectomas de fMRI	25
3 Resultados	28
3.1 Distancia espectral truncada	28
3.2 Wasserstein KDE	29
3.2.1 Wasserstein KDE muestral	31
3.2.2 Wasserstein KDE en malla	33
3.3 Gromov-Wasserstein	36
4 Conclusiones	39
Índice	41
Bibliografía	41

Agradecimientos

Agradezco a mis padres Marina y Yury por la paciencia y apoyo que me han dado toda la vida. Les agradezco la oportunidad a la educación, los valores que me han inculcado y la emoción por el estudio que me han transferido.

Gracias Said por tu apoyo incondicional, por escuchar mis ideas e impulsarme a seguir creciendo. Gracias por ser un gran equipo y festejar mis logros, a veces hasta más que yo.

Gracias a mi hermana y amiga, Nadia, quien, a pesar de tener muchas diferencias de opinión, siempre ha estado presente en cada etapa de mi carrera profesional y personal. Gracias a mis mejores amigas, Gabriela, Tania y Sam, por ser parte importante de mi vida y los momentos inolvidables que hemos pasado.

Un especial agradecimiento a mi director de tesis y sinodal Dr. Pablo Suárez Serato, por permitirme desarrollar un trabajo tan interesante, su apoyo y por todos los conocimientos que me ha compartido.

Otro especial agradecimiento al Dr. Pedro Miramontes Vidal, mi sinodal y tutor de la maestría, por ayudarme en alcanzar mis metas de graduarme de la maestría y seguir en el doctorado.

Por último, agradezco al resto el jurado, la Dra. Yalbi Itzel Balderas Martínez, el Dr. Marco Arieli Herrera Valdez y el Dr. Sergio Rajsbaum por su dedicación de tiempo e interés en mi trabajo.

Sin todos ustedes este trabajo no habría sido posible. ¡Un abrazo a todos!

Liuba

Introducción

El diagnóstico erróneo de enfermedades neurológicas es muy común. En efecto, existen muchos desórdenes neurológicos, pero una cantidad limitada de síntomas para determinar el desorden con exactitud. Los diagnósticos erróneos pueden provocar consecuencias severas en la salud del paciente, afectando su ocupación e interacciones sociales.

La proporción de diagnósticos incorrectos es, desafortunadamente, alta. Por ejemplo, cerca del 20% de nuevos casos de Esclerosis Múltiple son mal diagnosticados [1]. La mayoría de estos nuevos pacientes en realidad tenían migraña pero recibieron tratamientos innecesarios para esclerosis múltiple que duraron incluso años. Dichos tratamientos tienen efectos secundarios, como la leucoencefalopatía multifocal progresiva. Otro ejemplo es el síndrome de Tourette, confundido con la epilepsia, y hay muchos otros ejemplos más. Esto no es sorprendente, pues el cerebro es el órgano humano menos conocido.

Los ejemplos anteriores muestran la importancia de desarrollar métodos que puedan, de forma precisa y eficiente, confirmar un diagnóstico. Existe una comunidad muy grande de investigadores en distintas áreas que pretende entender el funcionamiento del cerebro. Una serie de investigaciones se hace sobre la actividad cerebral medida con Imágenes de Resonancia Magnética funcional (fMRI por sus siglas en inglés). En el artículo [2] analizan la actividad cerebral de pacientes en coma y personas control, y en la parte de discusión afirman que no hay diferencia evidente a nivel global y topológico entre las conexiones funcionales del cerebro entre los dos grupos de personas. La falta de entendimiento de la estructura topológica del cerebro es la motivación de este trabajo. Con ayuda de Topología, Estadística y Transporte Óptimo proponemos un algoritmo que puede ayudar a discriminar enfermedades neurológicas para ayudar a confirmar diagnósticos, rastrear la progresión de la enfermedad y monitorear los efectos terapéuticos.

En el primer capítulo definimos los conceptos de topología algebraica, teoría de grafos, estadística y transporte óptimo utilizados en este trabajo. Ahí mismo presentamos los ciclos sin propagación hacia atrás y la longitud espectral, y explicamos su importancia para cuantificar la diferencia entre conectomas, analizada anteriormente en [3]. Como veremos en este capítulo, la longitud espectral determina solo una parte del grafo: el subgrafo de nodos de grado mayor o igual a 2, ignorando el resto. Solucionamos este problema agregando al grafo un nodo adicional llamado cono, que hace que todos los nodos del grafo sean de grado mayor o igual a 2, ayudando así a determinar el grafo completo y evitando la pérdida de información.

En el siguiente capítulo describimos el conjunto de datos analizados en este traba-

jo, a saber, conectomas o grafos sin pesos no dirigidos que representan las conexiones funcionales entre regiones del cerebro. Los conectomas son obtenidos mediante el procesamiento de imágenes fMRI, siguiendo la metodología de [4], que describiremos a detalle en este mismo capítulo. El trabajo está limitado a las categorías de los conjuntos de datos que teníamos disponibles: personas en coma, personas control, jóvenes y viejos. Estos datos fueron proporcionadas por la investigadora Sophie Achard del “*Centre national de la recherche scientifique*”, autora de los artículos [2] y [4], mencionados anteriormente. Los conjuntos de datos otorgados promueven una investigación interesante, ya que resolvemos la falta de diferenciadores globales y topológicos entre los conectomas de distintas categorías, argumentada en la discusión de [2].

En el tercer capítulo describimos los algoritmos que cuantifican la diferencia topológica entre los conectomas, uniendo los conceptos definidos en el capítulo 1. Aquí mismo presentamos los resultados de estos algoritmos aplicados sobre los conjuntos de datos. También analizamos los resultados de agregar el nodo “cono” a los conectomas. Notamos que, dependiendo de la regla de conexión del cono a los demás nodos, el cono puede beneficiar el algoritmo y ayudar con la discriminación de conectomas o, al revés, perjudicarlo y hacer los conectomas menos distinguibles entre sí.

Concluimos la tesis con el capítulo 4, donde resumimos los resultados y exponemos el trabajo futuro.



Capítulo 1

Preliminares

1.1 Topología algebraica

La topología algebraica proporciona técnicas para formar estructuras algebraicas a partir de espacios topológicos. En general, estas estructuras algebraicas suelen ser grupos, pero aparecen estructuras más elaboradas como anillos, álgebras, etc. Esta sección revisa los conceptos de topología algebraica necesarios para este trabajo.

Definición 1. Una **trayectoria** en un espacio X es una función continua $f : I \rightarrow X$, donde $I = [0, 1]$.

Definición 2. Una **homotopía de trayectorias** en X es una familia $f_t : I \rightarrow X$, $0 \leq t \leq 1$ de trayectorias tal que

1. El punto inicial $f_t(0) = x_0$ y el punto final $f_t(1) = x_1$ son independientes de t .
2. La función $H : I \times I \rightarrow X$ definida por $H(s, t) = f_t(s)$ es continua.

Cuando dos trayectorias f_0 y f_1 están conectadas por una homotopía se dice que son homotópicas y se denota por $f_0 \simeq f_1$.

Ejemplo 3. Cualesquiera dos trayectorias f_0 y f_1 en \mathbb{R}^n con los mismos extremos x_0 y x_1 son homotópicas vía la homotopía $H(s, t) = (1 - t)f_0(s) + tf_1(s)$. En esta homotopía, cada punto $f_0(s)$ se une por una recta a $f_1(s)$, pues $f_0(s) + t(f_1(s) - f_0(s)) = (1 - t)f_0(s) + tf_1(s) = H(s, t)$. H en efecto es una homotopía, pues para $s = 0$ o $s = 1$, $H(0, t) = x_0$ y $H(1, t) = x_1 \forall t \in I$. La continuidad de la homotopía H como función de $I \times I \rightarrow \mathbb{R}^n$ se sigue de la continuidad de f_0 y f_1 , pues las operaciones de suma y multiplicación por escalar en la fórmula de H son continuas.

Proposición 4. *En cualquier espacio, la relación de homotopía entre trayectorias con extremos fijos es una relación de equivalencia.*

Demostración. La reflexividad es evidente puesto que $f \simeq f$ bajo la homotopía $H(s, t) = f(s)$. Para revisar la simetría, basta ver que si $f \simeq g$ bajo la homotopía $H(s, t)$, entonces $g \simeq f$ bajo la homotopía $H(s, 1 - t)$. Para demostrar la transitividad, supóngase

que $f \simeq g$ bajo H y $g \simeq h$ bajo G , con punto inicial x_0 y punto final x_1 . Entonces, $f \simeq h$ bajo la función

$$F(s, t) = \begin{cases} H(s, 2t) & 0 \leq t \leq \frac{1}{2} \\ G(s, 2t - 1) & \frac{1}{2} \leq t \leq 1 \end{cases}$$

F está bien definida pues para $t = \frac{1}{2}$, $H(s, 1) = g(s) = G(s, 0)$.

F es una homotopía, pues H y G son continuas y coinciden en $t = \frac{1}{2}$, por lo que F es continua. Además,

$$F(0, t) = \begin{cases} H(0, 2t) = x_0 & 0 \leq t \leq \frac{1}{2} \\ G(0, 2t - 1) = x_0 & \frac{1}{2} \leq t \leq 1 \end{cases}$$

Análogamente, $F(1, t) = x_1$. □

La clase de equivalencia de un trayectoria f se denota por $[f]$ y se conoce como clase de homotopía.

Definición 5. La **reparametrización** de un trayectoria f es la composición $f\phi$, donde $\phi : I \rightarrow I$ es una aplicación continua con $\phi(0) = 0$ y $\phi(1) = 1$.

La reparametrización respeta la clase de homotopía. En efecto, $f\phi \simeq f$ vía $f \circ H$, donde $H(s, t) = (1 - t)\phi(s) + ts$. Así, $H(s, 0) = \phi(s)$ y $H(s, 1) = s$. Nótese que $H(s, t) \in I$ para toda $(s, t) \in I \times I$, por lo que $f \circ H$ está bien definida.

Definición 6. Sean $f, g : I \rightarrow X$ dos trayectorias, con $f(1) = g(0)$. La **composición** o **el producto** de trayectorias $f \cdot g$ se define como

$$f \cdot g = \begin{cases} f(2s) & 0 \leq s \leq \frac{1}{2} \\ g(2s - 1) & \frac{1}{2} \leq s \leq 1 \end{cases}$$

El producto $f \cdot g$ es una trayectoria que atraviesa primero f y luego g . Esta operación respeta las clases de homotopía, pues, si $f_0 \simeq f_1$ vía la homotopía H y $g_0 \simeq g_1$ vía la homotopía G , y $f_0(1) = g_0(0)$, entonces $f_0 \cdot g_0 \simeq f_1 \cdot g_1$ vía la homotopía $F \cdot G$.

Definición 7. Las trayectorias $f : I \rightarrow X$ que tienen el mismo punto inicial y final, i.e. $f(0) = f(1) = x_0 \in X$, se llaman **ciclos** y x_0 se conoce como **punto base**. El conjunto de clases de homotopía $[f]$ de ciclos $f : I \rightarrow X$ con punto base x_0 se denota por $\pi_1(X, x_0)$.

Proposición 8. $\pi_1(X, x_0)$ con la operación producto $[f][g] = [f \cdot g]$ es un grupo.



Demostración. Al restringir el estudio de trayectorias a ciclos con punto base fijo $x_0 \in X$, garantizamos que la operación de producto $f \cdot g$ esté bien definida para cualesquiera dos ciclos f y g . Ya se ha hecho la observación que la operación producto respeta las clases de homotopía, por lo que $[f][g] = [f \cdot g]$ está bien definida. Dicho esto, lo que resta es demostrar los tres axiomas de un grupo, a saber, la asociatividad, la existencia del elemento neutro y el elemento inverso.

Sean f, g, h ciclos en X con punto base x_0 . Afirmamos que $f \cdot (g \cdot h)$ es una reparametrización de $(f \cdot g) \cdot h$. En efecto, sea ϕ la función lineal por pedazos definida como

$$\phi(s) = \begin{cases} 1/2s & 0 \leq s \leq \frac{1}{2} \\ s - 1 & \frac{1}{2} \leq s \leq \frac{3}{4} \\ 2s & \frac{3}{4} \leq s \leq 1 \end{cases}$$

Entonces, $f \cdot (g \cdot h) \circ \phi = f \cdot (g \cdot h)$, por lo que $f \cdot (g \cdot h) \simeq (f \cdot g) \cdot h$, demostrando que el producto en $\pi_1(X, x_0)$ es asociativo.

Sea $f : I \rightarrow X$ un ciclo con punto base x_0 , y sea $c : I \rightarrow X$ definida como $c(s) = x_0$, i.e. c es una trayectoria constante. Definamos las funciones $\phi_1, \phi_2 : I \rightarrow I$ como

$$\phi_1(s) = \begin{cases} 2s & 0 \leq s \leq \frac{1}{2} \\ 1 & \frac{1}{2} \leq s \leq 1 \end{cases} \quad \phi_2(s) = \begin{cases} 0 & 0 \leq s \leq \frac{1}{2} \\ 2s - 1 & \frac{1}{2} \leq s \leq 1 \end{cases}$$

Entonces, $f \cdot c = f\phi_1$ y $c \cdot f = f\phi_2$. Por lo tanto $c \cdot f \simeq f \simeq f \cdot c$. Así, hemos encontrado un elemento neutro en $\pi_1(X, x_0)$.

Por último, sea f un ciclo, y defínase \bar{f} el mismo ciclo pero con dirección inversa, esto es, $\bar{f} = f(1 - s)$. Entonces $f \cdot \bar{f} \simeq c$ vía la homotopía $h_t = f_t \cdot g_t$, donde f_t es igual a f en $[0, 1 - t]$ y a $f(1 - t)$ en $[1 - t, 1]$, y g_t es la inversa de f_t . Análogamente se demuestra que $\bar{f} \cdot f \simeq c$. Por lo tanto, $[f]$ es la inversa de $[f]$, demostrando que $\pi_1(X, x_0)$ es un grupo. \square

Este grupo se conoce como **grupo fundamental** de X en el punto base x_0 . Se verá más adelante que $\pi_1(X, x_0)$ es el primero de la secuencia de grupos $\pi_n(X, x_0)$, llamados grupos de homotopía, que son definidos de forma análoga sobre I^n en vez de I .

Ejemplo 9. Sea X un conjunto convexo de \mathbb{R}^n y consideremos un punto base $x_0 \in X$. El grupo fundamental $\pi_1(X, x_0) = 0$, es decir, X tiene grupo fundamental trivial. En efecto, cualquier ciclo f_0 es homotópico a $f = x_0$ vía la homotopía $f_t(s) = (1 - t)f_0(s) + tf_1(s)$.

Ya que $\pi_1(X, x_0)$ involucra solo una componente conectada por trayectorias de X , es claro que es posible encontrar una relación entre $\pi_1(X, x_0)$ y $\pi_1(X, x_1)$ solo cuando x_0 y x_1 se encuentran en la misma componente conectada por trayectorias.

Sean $h : I \rightarrow X$ una trayectoria de x_0 a x_1 , y \bar{h} la trayectoria inversa de h , con $\bar{h}(s) = h(1 - s)$. Si f es un ciclo basado en x_1 , entonces $h \cdot f \cdot \bar{h}$ es un ciclo basado en x_0 .



Esto define un cambio de punto base $\beta_h : \pi_1(X, x_1) \rightarrow \pi_1(X, x_0)$, con $\beta_h[f] = [h \cdot f \cdot \bar{h}]$. Esto está bien definido, pues si f_t es una homotopía de ciclos basados en x_1 , entonces $h \cdot f_t \cdot \bar{h}$ es una homotopía de ciclos basados en x_0 .

Proposición 10. *La función $\beta_h : \pi_1(X, x_1) \rightarrow \pi_1(X, x_0)$ definida arriba es un isomorfismo.*

Demostración. β_h es un homomorfismo, pues $\beta_h[f \cdot g] = [h \cdot f \cdot g \cdot \bar{h}] = [h \cdot f \cdot \bar{h} \cdot h \cdot g \cdot \bar{h}] = [h \cdot f \cdot \bar{h}][h \cdot g \cdot \bar{h}] = \beta_h[f]\beta_h[g]$.

β_h es un isomorfismo con inversa $\beta_{\bar{h}}$, pues $\beta_h\beta_{\bar{h}}[f] = \beta_h[\bar{h} \cdot f \cdot h] = [h \cdot \bar{h} \cdot f \cdot h \cdot \bar{h}] = [f]$. Análogamente sucede para $\beta_{\bar{h}}\beta_h$. \square

La proposición anterior demuestra que si X es conexo por trayectorias, el grupo $\pi_1(X, x_0)$ es, salvo isomorfismo, independiente de la elección del punto base. Por esto, la notación $\pi_1(X, x_0)$ se abrevia a $\pi_1(X)$. En este trabajo solo se consideran espacios conexos, por lo que adoptaremos la notación abreviada.

Definición 11. Las trayectorias cerradas sin un punto base distinguido son conocidas como **ciclos libres**.

Denotemos por I^n al cubo unitario de dimensión n , producto de n copias del intervalo $[0, 1]$. La frontera ∂I^n de I^n es un subespacio que consiste en los puntos donde al menos una de las coordenadas es 0 o 1.

Definición 12. Para un espacio X con punto base $x_0 \in X$, el **n -grupo fundamental** $\pi_n(X, x_0)$ es el conjunto de clases de homotopía de las funciones $f : (I^n, \partial I^n) \rightarrow (X, x_0)$, donde las homotopías f_t satisfacen $f_t(\partial I^n) = x_0$ para toda t .

Dados $f, g \in \pi_n(X, x_0)$, la operación producto se define como

$$f \cdot g(s_1, s_2, \dots, s_n) = \begin{cases} f(2s_1, s_2, \dots, s_n) & 0 \leq s_1 \leq \frac{1}{2} \\ g(2s_1 - 1, s_2, \dots, s_n) & \frac{1}{2} \leq s_1 \leq 1 \end{cases}$$

La definición se extiende a $n = 0$, tomando I^0 un punto, $\partial I^0 = \emptyset$, y $\pi_0(X, x_0)$ – el conjunto de las componentes por trayectorias de X .

Definición 13. Una **clase conjugada** del grupo fundamental de X es el conjunto C de clases de homotopía tal que para cualesquiera dos clases $[c_1], [c_2] \in C$, existe una tercer clase $[g]$ que satisface la igualdad $[c_1] = [\bar{g}][c_2][g]$, donde \bar{g} es la inversa de g .

Definición 14. Un **retracto por deformación** de un espacio X al subespacio A es una familia de funciones $f_t : X \rightarrow X$, $t \in I$, tal que $f_0 = \mathbb{I}$, donde \mathbb{I} es la función identidad, $f_1(X) = A$ y $f_t|_A = \mathbb{I}$ para toda t . Además se cumple que la función $F : X \times I \rightarrow X$ dada por $F(x, t) = f_t(x)$ es continua.

En la Figura 1.1 se pueden observar ejemplos de retracto por deformación.

Un retracto por deformación es un caso especial de la noción más general de homotopía entre funciones, definida más adelante.



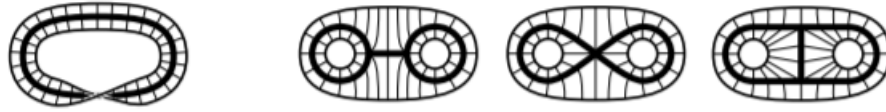


Figura 1.1: Ejemplos de retractos por deformación. La figura de la izquierda es un retracto por deformación de la banda de Möbius sobre un círculo. Las tres figuras de la derecha son retractos por deformación de un disco con dos bolas abiertas removidas que se puede encoger a tres subespacios distintos. La imagen fue extraída de [5]

Definición 15. Una **homotopía** es una familia de funciones $f_t : X \rightarrow Y$, $t \in I$ tal que la función asociada $F : X \times I \rightarrow Y$ dada por $F(x, t) = f_t(x)$ es continua. Dos funciones f_0 y f_1 son homotópicas si existe una homotopía f_t conectándolas.

Nótese que esta definición generaliza el concepto de homotopía de trayectorias.

Definición 16. Un **retracto** de X sobre A es una función $r : X \rightarrow X$ tal que $r(X) = A$ y $r|_A = \mathbb{I}$.

Definición 17. Un espacio X es **contráctil** si la función identidad $\mathbb{I} : X \rightarrow X$ es homótopa a una aplicación constante.

Definición 18. La **longitud espectral** \mathcal{L} , conocida también como **longitud espectral marcada**, es una función sobre $\pi_1(X)$ que asigna a cada clase de homotopía la longitud ínfima de todos los representantes de la clase conjugada. Nótese que la longitud de una clase de homotopía considera no solo la longitud de todos los ciclos de la clase, sino de todas sus clases conjugadas.

1.2 Teoría de grafos

Definición 19. Sean V un conjunto no vacío de vértices o nodos y E un conjunto de aristas (*pares de vértices*). Un **grafo** G es la estructura compuesta por (V, E) .

Ejemplo 20. Sean $V = \{0, 1, \dots, 15\}$ un conjunto de 16 nodos numerados del 0 al 15, y $E = \{(0, 1), (0, 2), \dots, (0, 15)\}$ las aristas que unen a los nodos $1, \dots, 15$ con el nodo 0. El grafo $G = (V, E)$ se puede visualizar de la siguiente manera:

Definición 21. Una **arista cíclica** es una arista que empieza y termina en el mismo vértice. Las **aristas múltiples** son aristas que se componen por la misma pareja de nodos, i.e. tienen mismo nodo inicial y mismo nodo final. Un **grafo simple** es un grafo que no tiene aristas cíclicas ni múltiples.

Cuando dos vértices son unidos por una arista, se dice que son **adyacentes** y que son **vecinos**.

Definición 22. Una **trayectoria** es un grafo simple cuyos vértices pueden ser ordenados de tal forma que dos vértices son adyacentes si y solo si son consecutivos en la lista.



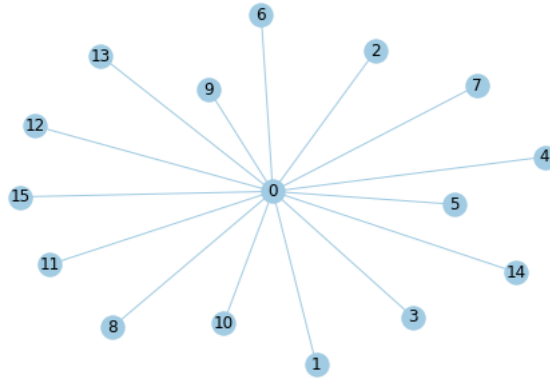


Figura 1.2: Ejemplo de grafo con estructura estrella

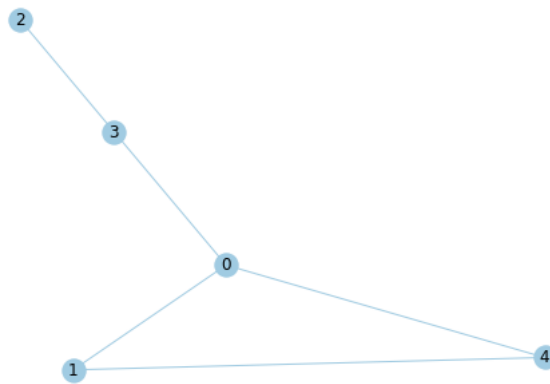


Figura 1.3: Ejemplo de grafo simple y conexo

Un **ciclo** es un grafo con la misma cantidad de nodos y de vértices, tal que los vértices se pueden acomodar en una lista circular con vértices adyacentes si y solo si aparecen de manera consecutiva en la lista circular. Un **triángulo** es un ciclo de 3 nodos.

Definición 23. Un **subgrafo** del grafo $G = (V_G, E_G)$ es un grafo $H = (V_H, E_H)$, con $V_H \subseteq V_G$ y $E_H \subseteq E_G$. Decimos que G contiene H y se escribe como $H \subseteq G$.

Un grafo G es **conexo** si cualesquiera dos vértices en G están conectados por una trayectoria. En caso contrario, G es **disconexo**.

Ejemplo 24. En la Figura 1.3 se muestra un grafo $G = (V, E)$, con $V = \{0, 1, 2, 3, 4\}$ y $E = \{(0, 1), (0, 4), (1, 4), (0, 3), (3, 2)\}$. Entonces G es un grafo simple conexo, que contiene un ciclo conformado por los vértices $\{0, 1, 4\}$. Este ciclo es también un triángulo.

Definición 25. Sea G un grafo con nodos $V = \{v_1, \dots, v_n\}$ y aristas $E = \{e_1, \dots, e_m\}$. La **matriz de adyacencia** de G es una matriz $A \in M_{n \times n}$, donde la entrada $a_{i,j}$ es la cantidad de aristas con extremos $\{v_i, v_j\}$.

Si el vértice v es un extremo de la arista e , entonces v y e son **incidentes**. El **grado** de un vértice v es el número de vértices incidentes.



Nota 26. La matriz de adyacencia siempre es simétrica. El grado de un nodo v es igual a la suma de las entradas del renglón (o columna) correspondiente a v de la matriz de adyacencia.

Ejemplo 27. En la Figura 1.3, el grado del vértice 0 es 3. Su matriz de adyacencia es

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (1.1)$$

Definición 28. Una **caminata** es una lista $v_0, e_0, v_1, \dots, e_k, v_{k+1}$, tal que para $0 \leq j \leq k$, los extremos de la arista e_j son v_j y v_{j+1} . La **longitud** de una trayectoria, un ciclo o una caminata es la cantidad de aristas que contiene.

Definición 29. Un **grafo dirigido** es un grafo que consta de un conjunto de vértices y un conjunto de aristas, tal que cada arista e se asocia a un par ordenado de vértices (v, w) , que se escribe como $e = (v, w)$ y se dice que la arista e va de v a w . Estas aristas se suelen representar como $v \rightarrow w$.

Definición 30. Un **grafo no dirigido** es un grafo $G = (V, E)$, donde cada arista $e \in E$ se asocia a un par de vértices no ordenado.

Definición 31. Un **grafo con pesos** es un grafo $G = (V, E)$, donde cada arista en E tiene asociado un peso, i.e. un número, un costo o un valor de cualquier tipo de dato.

Definición 32. Un grafo sin ciclos es **acíclico**. Un **bosque** es un grafo acíclico. Un **árbol** es un grafo acíclico conexo. Una **hoja** es un vértice de grado 1.

Ejemplo 33. Un árbol es un bosque conexo. Cada componente de un bosque es un árbol. Las trayectorias son árboles. Un árbol es una trayectoria si y solo si el grado máximo de sus nodos es dos.

La Figura 1.2 es un ejemplo de árbol, pues no tiene ciclos. Este árbol tiene una estructura particular donde un vértice es adyacente a todos los demás y es llamado estrella.

1.3 Ciclos sin retrocesos

En este trabajo nos interesa encontrar una forma de comparar grafos desde el punto de vista topológico. Para esto, se busca calcular la longitud espectral de las clases de homotopía del grafo. La motivación se basa en el teorema de Constantine y Lafont [6], donde se demuestra que la longitud espectral determina un subconjunto de la gráfica salvo isomorfismo, en particular la 2-núcleo subgráfica:

Teorema 34. *Sea X un espacio métrico. Se define $\text{Conv}(X)$ como el subconjunto mínimo al cual X se puede retraer por deformación. Sean X_1, X_2 dos espacios compactos no contraíbles, espacios geodésicos de dimensión topológica 1. Si la longitud espectral marcada de X_1, X_2 es la misma, entonces $\text{Conv}(X_1)$ es isométrico a $\text{Conv}(X_2)$*



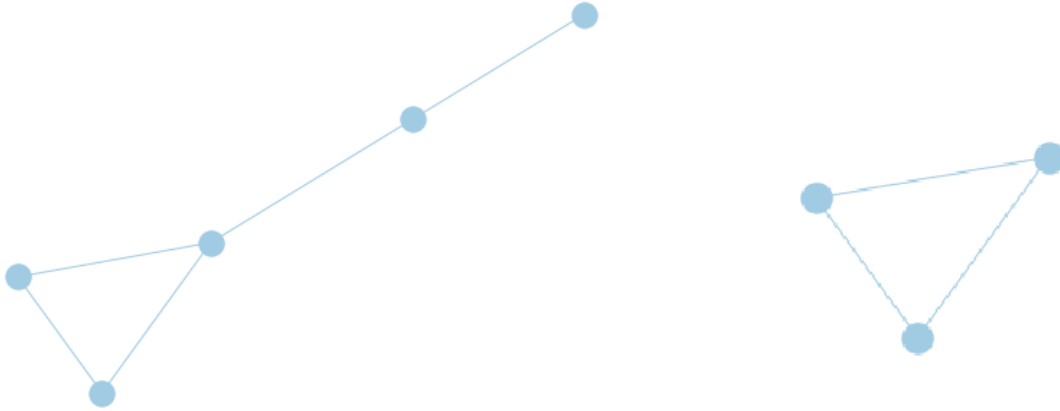


Figura 1.4: Del lado izquierdo se observa un grafo G de 5 nodos. Del lado derecho se observa $Conv(G)$

Por el teorema anterior, la similaridad de dos grafos se puede determinar comparando sus longitudes espectrales.

Si G es un grafo, $Conv(G)$ corresponde al subgrafo que resulta de remover iterativamente los nodos de grado 1, resultando en el subgrafo con nodos de grado mayor o igual a dos. Así, el teorema 34 dice que dos grafos G_1, G_2 con la misma longitud espectral tienen subgrafos $Conv(G_1)$ y $Conv(G_2)$ isomorfos.

Ejemplo 35. Sea G un grafo de 5 nodos como se muestra en la Figura 1.4. $Conv(G)$ es el subgrafo de 3 nodos del lado derecho de la Figura 1.4, que resulta de remover los nodos de grado 1.

Ejemplo 36. Si G es un árbol, entonces $Conv(G)$ es la retracción de G a un solo nodo.

El teorema 34 muestra una desventaja, y es que la longitud espectral determina solo la parte del grafo que tiene nodos de grado mayor o igual a 2, omitiendo todas las posibles variaciones del grafo con árboles anidados, haciéndolas indistinguibles entre sí. Por esto, proponemos agregar un cono al grafo analizado. El cono es un nodo adicional que se conecta a los demás nodos del grafo siguiendo una regla predefinida, forzando así a que todos los nodos del grafo tengan grado mayor o igual a 2. Con esto, a partir del grafo analizado, construimos un grafo X que cumple que $X = Conv(X)$. El teorema 34 dice que podemos determinar un grafo con cono dada su longitud espectral, evitando pérdida de información en el grafo original. Así, podemos comparar grafos con cono mediante la longitud espectral. En este trabajo analizaremos los cambios de la distancia entre conectomas al agregarles un cono. Usaremos dos reglas de conexión:

- 1) Cono conectado a todos los nodos de X .
- 2) Cono conectado a los nodos de X de grado 1.

El pseudocódigo del algoritmo de agregar un cono a un grafo se muestra a continuación.



Algorithm 1.3.1 $\text{add_cone}(\text{graph}, \text{method})$

Input: Un grafo y el método de conexión del cono a los demás nodos del grafo

Output: Grafo

```

1: procedure
2:    $\text{graph} \leftarrow$  un conectoma
3:    $\text{method} \leftarrow$  uno de los siguientes dos valores: all ó degree_one
4:   if  $\text{method} == \text{all}$  then
5:      $\text{graph} \leftarrow$   $\text{graph}$  con un nodo cono adicional
6:      $\text{graph} \leftarrow$   $\text{graph}$  con nodo cono conectado con todos los nodos del grafo
7:   end if
8:   if  $\text{method} == \text{degree\_one}$  then
9:      $\text{node\_list} \leftarrow$  lista vacía
10:    for  $\text{node}$  in  $\text{graph}$  do
11:      if  $\text{node.degree} == 1$  then
12:         $\text{node\_list} \leftarrow [\text{node\_list}, \text{node}]$ 
13:      end if
14:    end for
15:     $\text{graph} \leftarrow$   $\text{graph}$  con un nodo cono adicional
16:     $\text{graph} \leftarrow$   $\text{graph}$  con nodo cono conectado con  $\text{node\_list}$ 
17:  end if
18:  return  $\text{graph}$ 
19: end procedure

```

Por definición de longitud espectral, ésta se basa en el grupo fundamental del grafo. Todos los grafos que se analizan en este trabajo son grafos simples y conexos, por lo que se puede omitir el vértice específico que se usa como punto base para definir el grupo fundamental, ya que los grupos fundamentales coinciden.

Un concepto importante para el análisis del grupo fundamental de los grafos es el siguiente:

Definición 37. Sea G un grafo. Un **ciclo sin retroceso** en G es un ciclo en cuya secuencia las aristas no se atraviesan dos veces consecutivas, i.e. una arista no es seguida por su inversa. También nos referiremos a estos ciclos como **ciclos sin propagación hacia atrás**.

Ejemplo 38. Observemos de nuevo la Figura 1.3. Denotemos por $e_1 = (0, 1)$, $e_2 = (1, 4)$, $e_3 = (4, 0)$ las aristas del triángulo del grafo. Las inversas de estas aristas son $\bar{e}_1 = (1, 0)$, $\bar{e}_2 = (4, 1)$, $\bar{e}_3 = (0, 4)$. La secuencia $e_1 e_2 e_3$ forma un ciclo sin retrocesos, mientras que $e_1 e_2 e_2^{-1} e_2 e_3$ no lo es. Ahora, si denotamos a $e_4 = (0, 3)$, $\bar{e}_4 = (3, 0)$. Entonces la trayectoria $e_3 e_4 \bar{e}_4 e_1 e_2$ es un ciclo, pero no es un ciclo sin retrocesos.

Definición 39. La **matriz sin retrocesos** B es la matriz representante de todas las parejas de aristas que son secuencias sin retroceso de longitud 2 en el grafo $G = (V, U)$.



Si $m = |E|$ es la cantidad de aristas que tiene G , entonces la matriz B es una matriz cuadrada de $2m \times 2m$. Cada renglón y cada columna representa una arista $e \in E$ en una de sus dos direcciones. Para dos aristas $(u, v), (k, l) \in E$,

$$B_{k \rightarrow l, u \rightarrow v} = \delta_{vk}(1 - \delta_{ul})$$

donde δ_{ij} es la delta de Kronecker y $k \rightarrow l$ representa la dirección de la arista, en este caso del nodo k al nodo l .

Intuitivamente, la matriz B se puede interpretar como la matriz de transición de una caminata aleatoria que no realiza retrocesos.

Ejemplo 40. Sea G un árbol no dirigido de tres nodos y dos aristas como se muestra en la figura.



Por ser un grafo de dos aristas, la matriz sin retrocesos es de 4×4 y es la siguiente:

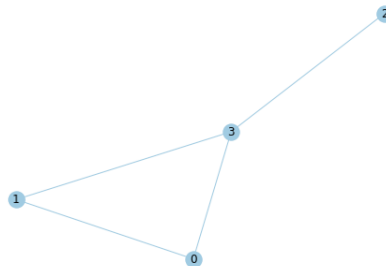
$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.2}$$

donde los renglones y columnas están acomodados en el orden $1 \rightarrow 2, 2 \rightarrow 3, 2 \rightarrow 1, 3 \rightarrow 2$. La entrada de valor 1 indica que las aristas forman una caminata de longitud dos sin retrocesos. Cuando es 0, significa que dicha secuencia de aristas no forma una caminata en el grafo, o que la caminata hace un retroceso. Vemos que en este ejemplo hay dos caminatas sin retroceso de longitud 2. Esto se puede observar directamente del grafo, donde las únicas caminatas sin retrocesos son: $1 \rightarrow 2 \rightarrow 3$ y $3 \rightarrow 2 \rightarrow 1$

Nótese que este ejemplo tiene caminatas sin retrocesos, pero no tiene ciclos sin retrocesos.

Ahora veamos un ejemplo que sí muestra ciclos sin retrocesos.

Ejemplo 41. Sea G un grafo simple, conexo, de cuatro nodos y cuatro aristas, como se muestra en la figura de abajo.



Al ser un grafo de 4 aristas, la matriz sin retrocesos será de 8×8 y es la siguiente:

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (1.3)$$

donde el orden de aristas es $0 \rightarrow 3, 0 \rightarrow 1, 1 \rightarrow 3, 2 \rightarrow 3, 3 \rightarrow 0, 1 \rightarrow 0, 3 \rightarrow 1, 3 \rightarrow 2$.

Contando las entradas de valor 1, vemos que el grafo de este ejemplo tiene 10 caminatas sin retrocesos de longitud dos.

1.3.1 Espectro de la matriz sin retrocesos

Los ciclos sin retrocesos de un grafo son importantes topológicamente, pues las aristas con retroceso, o los ciclos de longitud dos, son homotópicamente triviales, es decir, pueden ser contraídos a un punto. Por lo tanto, un ciclo sin retroceso es el representante de menor longitud de una clase de homotopía del grafo.

Como hemos visto, la matriz B representa aristas incidentes que no realizan retrocesos. Además de esto, en [3] se demuestra que la matriz B conlleva otra información útil, lo cual presentamos en el siguiente resultado.

Lema 42. $B_{k \rightarrow l, i \rightarrow j}^p$ equivale al número de trayectorias sin retroceso de longitud $p + 1$ que comienzan con la arista $i \rightarrow j$ y terminan en la arista $k \rightarrow l$. Más aún, $\text{tr}(B^p)$ es proporcional a la cantidad de ciclos sin retrocesos de longitud p en el grafo.

Demostración. Es evidente que la matriz B contiene la cantidad de caminos sin propagación hacia atrás de longitud 2. Sea M^k la matriz que contiene la cantidad de caminos sin propagación hacia atrás de longitud k , y M^{k+1} contiene a los de longitud $k + 1$. Note que

$$M_{u_1 \rightarrow u_2, v_1 \rightarrow v_2}^{k+1} = \sum_{w_1 \rightarrow w_2} M_{u_1 \rightarrow u_2, w_1 \rightarrow w_2}^k B_{w_1 \rightarrow w_2, v_1 \rightarrow v_2},$$

es decir,

$$M^{k+1} = M^k B.$$

Por inducción, $M^{k+1} = B^k$, lo que se quería demostrar.

Para la segunda parte, supongamos que $a \rightarrow b \rightarrow c \rightarrow a$ es un camino sin propagación hacia atrás, i.e., un triángulo. En este caso, $B_{c \rightarrow a, a \rightarrow b}^2$ es positiva, así como $B_{a \rightarrow b, a \rightarrow b}^3$. Lo último representa la existencia de un nodo c tal que el camino $a \rightarrow b \rightarrow c \rightarrow a \rightarrow b$ existe y no tiene propagaciones hacia atrás, por lo que $a \rightarrow b \rightarrow c \rightarrow a$ es un ciclo sin propagaciones hacia atrás. $B_{b \rightarrow c, b \rightarrow c}^3$ y $B_{c \rightarrow a, c \rightarrow a}^3$ también son positivas y también muestran la existencia del ciclo sin propagación hacia atrás a, b, c . Por lo tanto, la suma de los tres elementos de la diagonal $\text{tr}(B^3)$ cuenta cada triángulo exactamente seis veces, ya que hay tres aristas con dos orientaciones cada una. Por inducción se demuestra que lo mismo pasa para ciclos de cualquier longitud p . \square



Otro resultado importante de álgebra lineal, demostrado en [7], es el siguiente:

Lema 43. Sean $A \in \text{Mat}_{n \times n}(\mathbb{R})$ y $\lambda_1, \dots, \lambda_n$ los autovalores de A . Entonces

$$\text{tr}(A) = \sum_{j=1}^n \lambda_j \quad (1.4)$$

Demostración. Cada matriz A puede ser escrita de la forma $A = S^{-1}DS$, donde D es la matriz diagonal cuyas entradas son los valores propios de A , y S es la matriz formada por los vectores propios asociados. Entonces,

$$\text{tr}(A) = \text{tr}(S^{-1}DS) \stackrel{*}{=} \text{tr}(DSS^{-1}) = \text{tr}(D) = \sum_j \lambda_j$$

donde $*$ se sigue del hecho que $\text{tr}(A_1A_2) = \text{tr}(A_2A_1)$, donde A_1, A_2 son dos matrices cuadradas. \square

Los últimos dos resultados indican que la comprensión de la longitud espectral de un grafo se puede aproximar analizando los autovalores de su matriz sin retrocesos. Esto es muy útil, puesto que para encontrar la longitud espectral exacta se tendrían que encontrar todos los ciclos de cada clase de homotopía y sus conjuntos conjugados, los cuales son infinitos en un grafo y encontrarlos tiene complejidad exponencial.

El pseudocódigo del algoritmo para encontrar los eigenvalores de la matriz sin retrocesos de un grafo es el siguiente:

Algorithm 1.3.2 `nbeigs_calculate(graph, topk, fmt)`

Input: Un grafo, un entero que representa el número máximo de eigenvalores por calcular y el formato de los autovalores requerido

Output: Una lista de autovalores en el formato especificado

```

1: procedure
2:   graph  $\leftarrow$  un conectoma
3:   topk  $\leftarrow$  un entero
4:   fmt  $\leftarrow$  uno de los siguientes tres valores: 1D, 2D ó complex

5:   core  $\leftarrow$  graph sin los nodos de grado 1
6:   matrix  $\leftarrow$  matriz sin retrocesos de core
7:   eigs  $\leftarrow$  eigenvalores de matrix
8:   eigs  $\leftarrow$  eigenvalores ordenados por el tamaño de la norma, la parte real y la
   parte imaginaria
9:   eigs  $\leftarrow$  eigs en formato fmt
10: end procedure

```



1.4 Estadística

Definición 44. Sea $(\Omega, \mathcal{A}, P[\cdot])$ un espacio de probabilidad. Una **variable aleatoria** es una función $X : \Omega \rightarrow \mathbb{R}$. La función X tiene que cumplir que $A_r = \{\omega : X(\omega) \leq r\}$ sea un conjunto medible, i.e. $A_r \in \mathcal{A}$ para cualquier $r \in \mathbb{R}$.

Ejemplo 45. Considérese el experimento de lanzar una moneda. Sea X la función que vale 1 si la moneda cayó en cara y 0 si no. Entonces, tenemos $\Omega = \{\text{cara}, \text{sello}\}$, $\mathcal{A} = \{\emptyset, \{\text{cara}\}, \{\text{sello}\}, \Omega\}$ y $X(\omega) = 1$ si $\omega = \text{cara}$, y $X(\omega) = 0$ si $\omega = \text{sello}$. Así, la función X asocia un número real a cada evento que resulta de lanzar una moneda.

Demostremos que $A_r = \{\omega : X(\omega) \leq r\} \in \mathcal{A}$ para cualquier $r \in \mathbb{R}$. En efecto, si $r < 0$, entonces $\{\omega : X(\omega) \leq r\} = \emptyset \in \mathcal{A}$. Si $0 \leq r < 1$, entonces $\{\omega : X(\omega) \leq r\} = \{\text{sello}\} \in \mathcal{A}$, y si $r \geq 1$, entonces $\{\omega : X(\omega) \leq r\} = \Omega$. Por lo tanto, X es una variable aleatoria.

Definición 46. La **función de distribución** de una variable aleatoria X , denotada por $F_X(\cdot)$, es una función con dominio \mathbb{R} y contradominio $[0, 1]$ que satisface: $F_X(x) = P(X \leq x) = P[\{\omega : X(\omega) \leq x\}]$ para cualquier $x \in \mathbb{R}$.

Una función de distribución se define de manera única para cada variable aleatoria. Si se conoce, se puede usar para encontrar las probabilidades de eventos definidos en términos de su variable aleatoria correspondiente.

Ejemplo 47. Considere de nuevo el experimento de lanzar una moneda, y sea X la variable aleatoria que cuenta el número de caras. Entonces

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

Proposición 48. La función de distribución cumple las siguientes propiedades:

1. $F_X(-\infty) \equiv \lim_{x \rightarrow -\infty} F_X(x) = 0$ y $F_X(\infty) \equiv \lim_{x \rightarrow \infty} F_X(x) = 1$.
2. $F_X(\cdot)$ es monótona no decreciente, i.e. $F_X(a) \leq F_X(b)$ para $a < b$.
3. $F_X(\cdot)$ es continua por la derecha, i.e.

$$\lim_{0 < h \rightarrow 0} F_X(x+h) = F_X(x).$$

Demostración. 1. Sea $\{x_n\}$ una sucesión decreciente de números reales y $x_n \rightarrow -\infty$. Entonces, la sucesión de eventos $\{\omega : X(\omega) \leq x_n\}$ es una sucesión decreciente y

$$\bigcap_{n=1}^{\infty} \{\omega : X(\omega) \leq x_n\} = \emptyset.$$

Por lo tanto

$$\lim_{n \rightarrow -\infty} F(x_n) = \lim_{n \rightarrow -\infty} P(\{\omega : X(\omega) \leq x_n\}) = P(\emptyset) = 0$$



Del mismo modo, si $\{x_n\}$ es una sucesión creciente y $x_n \rightarrow \infty$, entonces la sucesión de eventos $\{\omega : X(\omega) \leq x_n\}$ es una sucesión creciente y

$$\cup_{n=1}^{\infty} \{\omega : X(\omega) \leq x_n\} = \Omega.$$

Por lo tanto

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(\{\omega : X(\omega) \leq x_n\}) = P(\Omega) = 0.$$

2. Si $x_1 < x_2$, entonces

$$F(x_2) - F(x_1) = P(X \leq x_2) - P(X \leq x_1) = P(x_1 < X \leq x_2) \geq 0$$

Por lo tanto $F(x)$ es una función no decreciente.

3. Para probar que F es continua por la derecha, basta demostrar que si $\{x_n\}$ es una sucesión decreciente que tiende a a , entonces

$$\lim_{n \rightarrow \infty} F(x_n) = F(a)$$

Ya que

$$\{X \leq a\} = \cup_n \{X \leq x_n\}$$

y $\{X \leq x_n\}$ es una sucesión decreciente, entonces

$$\lim_n F(x_n) = \lim_n P(X \leq x_n) = P(X \leq a) = F(a)$$

□

Definición 49. Toda función que cumpla las propiedades de la proposición anterior es una **función de distribución**.

Se puede demostrar que ambas definiciones de función de distribución son equivalentes, véase [8].

Definición 50. Una variable aleatoria X es **discreta** si su rango es numerable, y es **continua** si existe una función $f_X(\cdot)$ tal que $F_X(x) = \int_{-\infty}^x f_X(u) du$ para cualquier número real x .

Definición 51. Si X es una variable aleatoria discreta con valores $x_1, x_2, \dots, x_n, \dots$, entonces la función

$$f_X(x) = \begin{cases} P[X = x_j] & \text{si } x = x_j, \quad j = 1, 2, \dots \\ 0 & \text{en cualquier otro caso} \end{cases}$$

es la **función de densidad discreta** de X . Si X es una variable aleatoria continua, entonces $f_X(\cdot)$ en $F_X(x) = \int_{-\infty}^x f_X(u) du$ es la **función de densidad continua** de X .

Teorema 52. Sea X una variable aleatoria (continua o discreta). Entonces, la función de distribución $F_X(\cdot)$ puede ser obtenida de la función de densidad $f_X(\cdot)$ y viceversa.



Demostración. Consideremos primero el caso discreto. Sean x_1, x_2, \dots las masas de X , i.e. x_j tal que $P[X = x_j] \neq 0$, para $j = 1, 2, \dots$. Si $f_X(\cdot)$ está dada, entonces

$$F_X(x) = \sum_{j: x_j \leq x} f_X(x_j).$$

Si $F_X(\cdot)$ está dada, entonces

$$f_X(x_j) = F_X(x_j) - \lim_{0 < h \rightarrow 0} F_X(x_j - h),$$

por lo que $f_X(x_j)$ puede ser encontrada para cualquier masa x_j de X . Definiendo $f_X(x) = 0$ para $x \neq x_j$, obtenemos la función de densidad discreta $f_X(x)$.

Para el caso continuo, si $f_X(\cdot)$ está dada, tenemos por definición que

$$F_X(x) = \int_{-\infty}^x f_X(u) du.$$

Por otro lado, si $F_X(\cdot)$ es conocida, entonces $f_X(x)$ se puede obtener diferenciando la función de distribución, i.e. $f_X(x) = \frac{dF_X(x)}{dx}$ en puntos x donde $F_X(x)$ es diferenciable. \square

A continuación, veremos las distribuciones más conocidas.

Definición 53. Una variable aleatoria X es una **variable aleatoria discreta uniforme** si su función de densidad es de la forma

$$f(x) = \left\{ \begin{array}{ll} \frac{1}{N} & x = 1, 2, \dots, N \\ 0 & \end{array} \right\} = \frac{1}{N} I_{\{1, 2, \dots, N\}}(x)$$

donde I es la función indicadora y $N \in \mathbb{N}$.

Definición 54. Si la función de densidad de una variable aleatoria X está dada por

$$f_X(x) = f_X(x; a, b) = \frac{1}{b - a} I_{[a, b]}(x),$$

$-\infty < a < b < \infty$, entonces la variable aleatoria es **uniforme continua**, distribuida sobre $[a, b]$.

Definición 55. Una variable aleatoria X tiene **distribución normal con media μ y varianza σ^2** si su función de densidad está dada por

$$f_X(x) = f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2},$$

con $-\infty < \mu < \infty$ y $\sigma > 0$.

Conocer la función de densidad de una variable aleatoria es de gran utilidad pues describe el comportamiento de dicha variable. En general, la función de densidad no es conocida, sin embargo existen técnicas para estimarla.



1.4.1 Estimación de densidad con núcleo

La estimación de densidad con núcleo (KDE, por sus siglas en inglés) es uno de los métodos de estimación más conocidos. KDE es un estimador no paramétrico, i.e. no requiere del supuesto de que la distribución buscada pertenezca a una familia de distribuciones conocidas, como la normal o la uniforme, haciendo a KDE una técnica muy flexible y popular, sobre todo para datos que resultan tener una distribución complicada.

Sean $X_1, \dots, X_n \in \mathbb{R}^d$ variables aleatorias independientes, idénticamente distribuidas provenientes de una distribución desconocida P con densidad p . Formalmente, KDE es expresada como

$$\hat{p}_h(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

donde $K : \mathbb{R}^d \rightarrow \mathbb{R}$ es una función suave llamada núcleo, y $h > 0$ es el ancho de ventana que controla la suavidad. Ejemplos comunes de $K(x)$ son los siguientes

$$\text{Núcleo Gaussiano: } K(x) = \frac{\exp(-\|x\|^2/2)}{v_{1,d}}, \quad v_{1,d} = \int \exp(-\|x\|^2/2) dx$$

$$\text{Núcleo Esférico: } K(x) = \frac{I(\|x\| \leq 1)}{v_{2,d}}, \quad v_{2,d} = \int I(\|x\| \leq 1) dx$$

Intuitivamente, KDE tiene el efecto de expandir cada punto X_j , con $j = 1, \dots, n$ en una función flar suave, cuya forma es determinada por el núcleo $K(x)$. Después, se suman las funciones flar resultantes, obteniendo el estimador de densidad final. En regiones con muchas observaciones, al tener muchas funciones flar por sumarse, KDE arrojará un valor alto. Por otro lado, si una región tiene pocas observaciones, la densidad estimada será baja pues la suma de las funciones flar estará compuesta por relativamente pocos términos, como se muestra en la Figura 1.5.

Para más sobre KDE, recomiendo el tutorial de Yen-Chi Chen [9].

1.5 Transporte Óptimo

La teoría de transporte óptimo es el estudio de la transportación y distribución óptima de recursos. Es un área matemática con una amplia variedad de aplicaciones, tales como segmentación y restauración de imágenes, análisis de datos, visión computacional y cualquier área que conlleve distribuciones de densidad.

En esta sección daremos una descripción muy breve de esta teoría, limitándonos solo a los conceptos y funciones necesarios para este trabajo. Invitamos al lector interesado a revisar [10] para profundizar la parte teórica, y consultar [11] para la parte práctica de transporte óptimo.

El desarrollo de la teoría se remonta a Gaspard Monge (1746-1818), cuando expone el problema de mover, con un mínimo esfuerzo, una pila de arena de gran tamaño (*cuya forma se puede considerar como una distribución de probabilidad*) en la forma de otra pila de arena de forma prescrita [12]. En otras palabras, se busca encontrar la



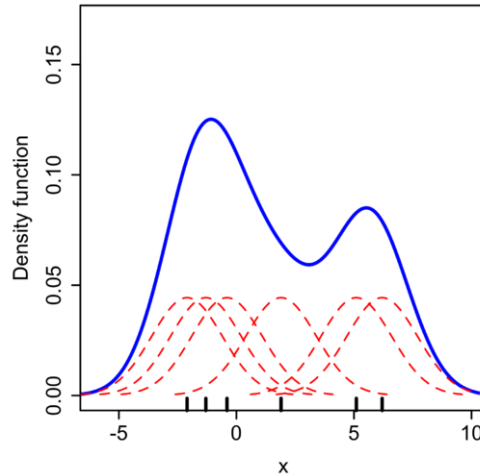


Figura 1.5: Las líneas negras en el eje x son las observaciones cuya densidad se quiere estimar. Las líneas punteadas rojas son las funciones flan que se generan y posteriormente se suman, obteniendo así la función de densidad estimada que está de color azul.

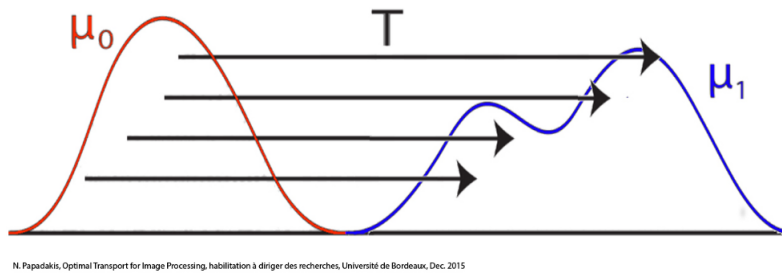


Figura 1.6: Transformación T que mueve la densidad μ_0 en la densidad μ_1

transformación óptima T para mover un conjunto de objetos (distribuciones discretas) de un lugar a otro. De manera similar para el caso continuo, se busca la transformación óptima T para mover una función de densidad en otra, como se muestra en la Figura 1.6.

La teoría de transporte óptimo está desarrollada tanto para el caso discreto como para el continuo. Daremos las definiciones de ambos casos. El caso continuo se separará mediante cuadros grises.

Definición 56. El **vector de probabilidad o histograma** \mathbf{a} es un vector de dimensión finita cuyas entradas son no negativas y suman 1:

$$\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}_+^n : \sum_{i=1}^n a_i = 1 \tag{1.5}$$

Definición 57. El **simplejo de probabilidad** Σ_n es el conjunto de todos los vectores de probabilidad. Matemáticamente se escribe así:

$$\Sigma_n := \{ \mathbf{a} \in \mathbb{R}_+^n : \sum_{i=1}^n a_i = 1 \} \tag{1.6}$$



Definición 58. Una **medida discreta** α con pesos \mathbf{a} y ubicaciones $x_1, \dots, x_n \in X$ se escribe de la forma:

$$\alpha = \sum \mathbf{a}_i \delta_{x_i} \tag{1.7}$$

donde δ_x es la función de Dirac en x . Si \mathbf{a} es un vector de probabilidad, entonces α es una **medida de probabilidad**.

Intuitivamente, la medida discreta representa puntos con pesos en el espacio.

Definición 59. Una **medida arbitraria** α en el espacio métrico (X, d) se puede evaluar al integrarla con cualquier función continua f :

$$\int_{\mathcal{X}} f(x) d\alpha(x) \tag{1.8}$$

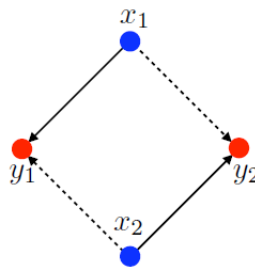
Se consideran solo las medidas de Radon $\mathcal{M}(\mathcal{X})$, i.e. medidas cuya integral de la ecuación 1.8 es finita.

Definición 60. El problema más sencillo de transporte óptimo es el **problema de asignación**. Sean $a, b \in \mathbb{R}^n$ dos vectores de probabilidad con distribución uniforme, i.e. $\mathbf{a}=\mathbf{b}=(\frac{1}{n}, \dots, \frac{1}{n})$. Dada una matriz de costo $(C_{i,j})$, se busca una permutación σ que minimice el costo de transportación de \mathbf{a} a \mathbf{b} , es decir, se busca resolver

$$\min_{\sigma \in \text{Perm}(n)} \frac{1}{n} \sum_{i=1}^n C_{i, \sigma(i)}$$

Note que la solución al problema de asignación no es única. Veamos esto en el siguiente ejemplo.

Ejemplo 61. Considérese el problema de transportar $x_i \rightarrow y_i$ para $i, j \in \{1, 2\}$, como se muestra en la siguiente figura.



El costo de transporte de $x_i \rightarrow y_i$ está dado por la matriz (C_{ij})

$$C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

En este caso solo hay dos transformaciones posibles:

$$T_1(x_1) = y_1, T_1(x_2) = y_2 \quad \text{o} \quad T_2(x_1) = y_2, T_2(x_2) = y_1$$

y el costo de transporte de ambas transformaciones es 2. Por ende, la solución al problema de asignación no es única.



El problema de asignación es el ejemplo más básico de transporte óptimo y se generaliza a vectores de probabilidad no uniformes, medidas discretas y continuas como se verá a continuación.

Definición 62. Sean $\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^m$ dos medidas discretas

$$\alpha = \sum a_i \delta_{x_i} \quad \beta = \sum b_i \delta_{y_i}$$

y sea $c : \{x_i\}_{i=1}^n \rightarrow \{y_j\}_{j=1}^m$ una función de costo. Al problema de encontrar la transformación óptima $T : \{x_1, \dots, x_n\} \rightarrow \{y_1, \dots, y_m\}$ tal que minimice el costo de transportación y que cumpla la conservación de masa

$$b_j = \sum_{i:T(x_i)=y_j} a_i$$

se le conoce como **problema de Monge**.

El operador de conservación de masa se llama operador **push forward** y se denota por

$$T_{\#}\alpha = \beta$$

Dicho esto, el problema de Monge se escribe como

$$\min_T \left\{ \sum_i c(x_i, T(x_i)) : T_{\#}(\alpha) = \beta \right\}$$

Ejemplo 63. Sea $\mathbf{a}=(\frac{1}{2}, 1)$ y $\mathbf{b}=\frac{3}{2}$. Las medidas discretas formadas por estos vectores son $\alpha = a_1\delta_{x_1} + a_2\delta_{x_2}$ y $\beta = b\delta_{y_2}$. Tomemos como función de costo una función $c : \{x_1, x_2\} \rightarrow y_2$ cualquiera. En este ejemplo solo existe una transformación posible, transportando los pesos de ambas posiciones x_1, x_2 a y_2 , Esta transformación cumple con la conservación de masa, pues $a_1 + a_2 = b_1$, y es la transformación óptima (y única) para este ejemplo. Véase imagen 1.7 para visualización.

Sean $\alpha \in \mathcal{M}(\mathcal{X}), \beta \in \mathcal{M}(\mathcal{Y})$ una medida de Radon. El **operador push-forward** para el caso medible se define como $T_{\#}(\alpha) = \beta$ si se cumple

$$\forall h \in \mathcal{C}(\mathcal{Y}) \quad \int_{\mathcal{Y}} h(y) d\beta(y) = \int_{\mathcal{X}} h(T(x)) d\alpha(x)$$

o, equivalentemente, para cualquier conjunto medible $B \in \mathcal{Y}$

$$\beta(B) = \alpha(T^{-1}(B))$$

El **problema de Monge** para medidas arbitrarias se define como

$$\min_T \left\{ \int c(x, T(x)) d\alpha(x) : T_{\#}(\alpha) = \beta \right\}$$

donde $c(x, y) : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ es la función de costo.



En el mundo real no siempre se cumple la conservación de masa. Por ejemplo, si queremos transportar productos desde bodegas a sucursales, no siempre se va a cumplir que una sucursal va a tener espacio suficiente para guardar todos los productos de una o más bodegas. Además de esto, en general se necesita distribuir productos de una bodega a varias sucursales distintas, mientras que en el problema de Monge se asume que se transporta todo de una bodega a un mismo destino. Esta restricción causa que el problema de Monge no sea simétrico: en el ejemplo anterior existe una transformación óptima que transporta la medida α en β , pero no existe una transformación de β a α .

Kantorovich contribuyó a la relajación de estas restricciones, y redefinió el problema de Monge para hacerlo simétrico.

1.5.1 Relajación de Kantorovich

La idea clave de Kantorovich es cambiar la naturaleza determinista del transporte al relajar el hecho de que la masa de un punto fuente x_i solo puede asignarse a un punto único $y_j = T(x_i)$. Kantorovich propone que la masa de cualquier punto x_i se pueda distribuir en varias ubicaciones. De esta manera Kantorovich se aleja de la idea del transporte determinista para considerar el transporte probabilístico, permitiendo la división de masa hacia varios destinos [13]. Esta distribución de masas se codifica con la matriz de acoplamiento.

Definición 64. Sean $\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^m$ dos medidas discretas

$$\alpha = \sum a_i \delta_{x_i}, \quad \beta = \sum b_i \delta_{y_i}$$

Una **matriz de acoplamiento** es una matriz $P \in \mathbb{R}^{n \times m}$, donde la entrada P_{ij} describe la cantidad de masa que fluye de a_i a b_j . Una matriz de acoplamiento P es **admisibles** si la suma de las masas transportadas a cierta ubicación coincide con la masa total de dicha ubicación.

El conjunto de matrices de acoplamiento admisibles se denota por

$$U(a, b) := \{P \in \mathbb{R}_+^{n \times m} : P\mathbf{1}_m = a, \quad P^T\mathbf{1}_n = b\},$$

donde $\mathbf{1}_m$ es un vector de unos de dimensión m .

El **problema de transporte óptimo de Kantorovich** se escribe como:

$$L_C(a, b) := \min_{P \in U(a, b)} \langle C, P \rangle$$

En vez de buscar una transformación óptima T para la distribución de recursos, ahora se busca la *matriz de acoplamiento* óptima $P \in \mathbb{R}^{n \times m}$. Con la relajación de Kantorovich el problema de transportar recursos se vuelve un problema simétrico, en el sentido de que si P es una matriz de acoplamiento óptima admisible para transportar la medida discreta α en β , entonces P^T es una matriz de acoplamiento óptima admisible para transportar β en α . Matemáticamente, si $P \in U(a, b)$, entonces $P^T \in U(b, a)$.



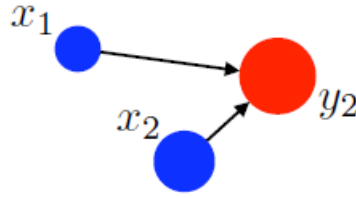


Figura 1.7: La suma de las masas de los puntos azules equivale a la masa del punto rojo

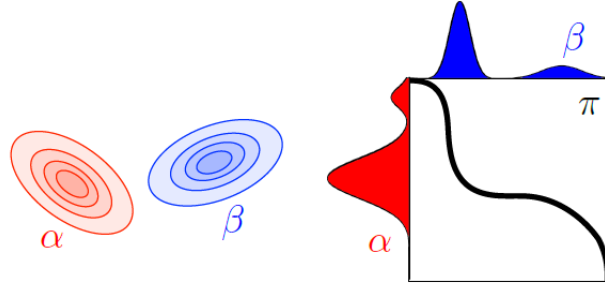


Figura 1.8: Función de acoplamiento π entre medidas de Radon α y β . Imagen extraída del libro de Transporte Óptimo Computacional de Gabriel Peyré y Marco Cuturi [11]

Definición 65. Las matrices de acoplamiento se generalizan a **funciones de acoplamiento** para el caso continuo como sigue:

$$\mathcal{U}(\alpha, \beta) := \{ \pi \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y}) : P_{\mathcal{X}\#} = \alpha, \quad P_{\mathcal{Y}\#} = \beta \}$$

$$\text{donde } P_{\mathcal{X}}(x, y) = x \quad \text{y} \quad P_{\mathcal{Y}}(x, y) = y$$

Véase el ejemplo visual en Figura 1.8.

Definición 66. El **problema de transporte óptimo de Kantorovich** para medidas arbitrarias se escribe como:

$$\mathcal{L}_c(\alpha, \beta) := \min_{\pi \in \mathcal{U}(\alpha, \beta)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y)$$

Si la matriz de costo C proviene de una distancia D , entonces el problema de transporte óptimo de Kantorovich resulta ser una distancia entre histogramas y medidas de probabilidad.

Proposición 67. Si la matriz de costo $C = D^p = (D_{i,j}^p)_{i,j} \in \mathbb{R}^{n \times n}$, donde $D \in \mathbb{R}_+^{n \times n}$ es una distancia, i.e.

1. $D \in \mathbb{R}_+^{n \times n}$ es simétrica
2. $D_{i,j} = 0$ si y solo si $i = j$
3. $\forall (i, j, k) \in [[n]]^3, D_{i,k} \leq D_{i,j} + D_{j,k}$,



entonces

$$W_p(\mathbf{a}, \mathbf{b}) := L_{D^p}(\mathbf{a}, \mathbf{b})^{1/p}$$

es una distancia en Σ_n , i.e. W_p es simétrica, positiva, $W_p(\mathbf{a}, \mathbf{b}) = 0 \iff \mathbf{a} = \mathbf{b}$, y satisface la desigualdad del triángulo:

$$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \Sigma_n, \quad W_p(\mathbf{a}, \mathbf{c}) \leq W_p(\mathbf{a}, \mathbf{b}) + W_p(\mathbf{b}, \mathbf{c})$$

Demostración. Ya que $C = D^p$ tiene diagonal nula, entonces $W_p(\mathbf{a}, \mathbf{a}) = 0$, con la matriz de acoplamiento óptima $P^* = \text{diag}(\mathbf{a})$. Ya que, a excepción de la diagonal, todos los elementos en D^p son positivos, entonces $W_p(\mathbf{a}, \mathbf{b}) > 0$ cuando $\mathbf{a} \neq \mathbf{b}$, pues así la matriz de acoplamiento admisible necesariamente tendrá una entrada positiva fuera de la diagonal. Como D^p es simétrica, $W_p(\mathbf{a}, \mathbf{b})$ lo es también.

Para probar la desigualdad del triángulo, sean $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \Sigma_n$ y P, Q soluciones óptimas al problema de transporte entre \mathbf{a} y \mathbf{b} , y entre \mathbf{b} y \mathbf{c} , respectivamente. Para evitar la división entre 0, definamos el vector $\tilde{\mathbf{b}}$ tal que

$$\tilde{b}_j = \begin{cases} b_j & \text{si } b_j > 0 \\ 1 & \text{si } b_j = 0 \end{cases}$$

Con base en esto, definamos la matriz S como sigue:

$$S := P \text{diag}(1/\tilde{\mathbf{b}}) Q \in \mathbb{R}_+^{n \times n}.$$

Note que $S \in U(\mathbf{a}, \mathbf{c})$, pues

$$S \mathbf{1}_n = P \text{diag}(1/\tilde{\mathbf{b}}) Q \mathbf{1}_n = P(\mathbf{b}/\tilde{\mathbf{b}}) = P \mathbf{1}_{\text{Supp}(\mathbf{b})} = \mathbf{a}$$

donde $\mathbf{1}_{\text{Supp}(\mathbf{b})}$ es el vector de dimensión n con 1s en la entrada j donde $b_j > 0$, y 0s en las demás entradas. La última igualdad se verifica porque $P \mathbf{1}_{\text{Supp}(\mathbf{b})} = P \mathbf{1} = \mathbf{a}$, pues $P_{i,j} = 0$ para j tales que $b_j = 0$. De la misma manera se muestra que $S^T \mathbf{1}_n = \mathbf{c}$. Así,

$$\begin{aligned} W_p(\mathbf{a}, \mathbf{c}) &= \left(\min_{P \in U(\mathbf{a}, \mathbf{c})} \langle P, D^p \rangle \right)^{1/p} \leq \langle S, D^p \rangle^{1/p} \\ &= \left(\sum_{ik} D_{ik}^p \sum_j \frac{P_{ij} Q_{jk}}{\tilde{b}_j} \right)^{1/p} \leq \left(\sum_{ijk} (D_{ij} + D_{jk})^p \frac{P_{ij} Q_{jk}}{\tilde{b}_j} \right)^{1/p} \\ &\leq \left(\sum_{ijk} D_{ij}^p \frac{P_{ij} Q_{jk}}{\tilde{b}_j} \right)^{1/p} + \left(\sum_{ijk} D_{jk}^p \frac{P_{ij} Q_{jk}}{\tilde{b}_j} \right)^{1/p} \end{aligned}$$

La primer desigualdad se cumple por la suboptimalidad de S , la segunda por la desigualdad del triángulo de D , y la tercer desigualdad se sigue de la desigualdad de Minkowski.

Así,

$$\begin{aligned} W_p(\mathbf{a}, \mathbf{b}) &\leq \left(\sum_{ij} D_{ij}^p P_{ij} \sum_k \frac{Q_{jk}}{\tilde{b}_j} \right)^{1/p} + \left(\sum_{jk} D_{jk}^p Q_{jk} \sum_i \frac{P_{ij}}{\tilde{b}_j} \right)^{1/p} \\ &= \left(\sum_{ij} D_{ij}^p P_{ij} \right)^{1/p} + \left(\sum_{jk} D_{jk}^p Q_{jk} \right)^{1/p} \\ &= W_p(\mathbf{a}, \mathbf{b}) + W_p(\mathbf{b}, \mathbf{c}) \end{aligned}$$



demostrando la desigualdad del triángulo. \square

Definición 68. La distancia de la proposición anterior es conocida como **distancia de Wasserstein**

Proposición 69. Si $\mathcal{X} = \mathcal{Y}$ y para alguna $p \geq 1$ se tiene que $c(x, y) = d(x, y)^p$, donde d es una distancia en \mathcal{X} , i.e.

1. $d(x, y) = d(y, x) \geq 0$;
2. $d(x, y) = 0 \iff x = y$;
3. $\forall (x, y, z) \in \mathcal{X}^3, d(x, z) \leq d(x, y) + d(y, z)$

entonces la **p -distancia de Wasserstein** entre medidas arbitrarias:

$$\mathcal{W}_p(\alpha, \beta) = \mathcal{L}_{d^p}(\alpha, \beta)^{1/p} \tag{1.9}$$

es, en efecto, una distancia, i.e. \mathcal{W}_p es simétrica, no negativa, $\mathcal{W}_p(\alpha, \beta) = 0 \iff \alpha = \beta$, y satisface la desigualdad del triángulo:

$$\forall (\alpha, \beta, \gamma) \in \mathcal{M}_+^1(\mathcal{X})^3, \quad \mathcal{W}_p(\alpha, \gamma) \leq \mathcal{W}_p(\alpha, \beta) + \mathcal{W}_p(\alpha, \beta)$$

Demostración. Se demuestra de la misma manera que la proposición 67: se construye un acoplamiento entre (α, γ) usando los acoplamientos óptimos de (α, β) y (β, γ) . \square

El problema de transporte óptimo es un problema de optimización que tiene como objetivo minimizar la función de costo de transporte.

1.5.2 Distancia de Gromov-Wasserstein

Hasta ahora hemos asumido la existencia de una función de costo C que cuantifica el costo de transportar histogramas (\mathbf{a}, \mathbf{b}) o medidas discretas (α, β) . La teoría de transporte óptimo no puede ser aplicada si no hay una función de costo definida entre dichos objetos. Esto sucede cuando los histogramas o medidas no están definidas en el mismo espacio.. Para abordar dicha restricción, se hace una suposición débil: existen dos matrices $D \in \mathbb{R}^{n \times n}$ y $D' \in \mathbb{R}^{m \times m}$ que cuantifican la similaridad entre las entradas de cada uno de los histogramas y/o medidas. Un caso de interés es cuando D y D' provienen de matrices de distancia, dando lugar a la distancia de Gromov-Wasserstein, introducida y estudiada por Facundo Mémoli [14].

Definición 70. Sean $a \in \mathbb{R}^n$ y $b \in \mathbb{R}^m$ dos histogramas. Sea $D \in \mathbb{R}^{n \times n}$ la matriz de distancia entre las entradas de a y $D' \in \mathbb{R}^{m \times m}$ la matriz de distancias entre las entradas del vector b .

Entonces, la **distancia de Gromov-Wasserstein** entre a y b es



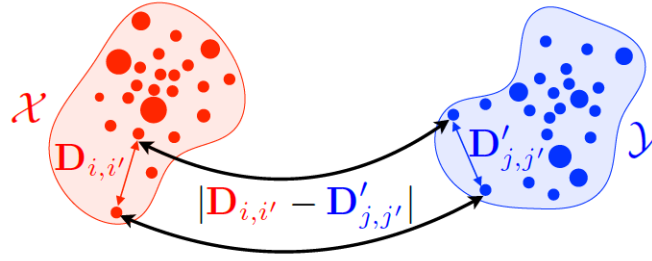


Figura 1.9: Imagen extraída del libro de Transporte Óptimo Computacional de Gabriel Peyré y Marco Cuturi [11]

$$GW((a, D), (b, D'))^2 = \min_{P \in U(a,b)} \mathcal{E}_{D,D'}(P)$$

donde

$$\mathcal{E}_{D,D'}(P) = \sum_{i,j,i',j'} |D_{i,i'} - D'_{j,j'}|^2 P_{i,j} P_{i',j'}$$

Si α y β son medidas discretas con pesos a y b , respectivamente, y D, D' son las matrices de distancia definidas anteriormente, entonces la distancia de Gromov Wasserstein entre α y β se define como $GW((\alpha, D), (\beta, D')) = GW((a, D), (b, D'))$.

La distancia de Gromov-Wasserstein busca mapear los puntos a en b con el mínimo ajuste posible, así como se muestra en la Figura 1.9.

Sean $(\mathcal{X}, d_{\mathcal{X}}, \alpha_{\mathcal{X}})$ y $(\mathcal{Y}, d_{\mathcal{Y}}, \alpha_{\mathcal{Y}})$ dos espacios métricos, donde $(d_{\mathcal{X}}, d_{\mathcal{Y}})$ son distancias, y $\alpha_{\mathcal{X}}, \alpha_{\mathcal{Y}}$ son medidas. La distancia de Gromov-Wasserstein

$$\mathcal{GW}((\alpha_{\mathcal{X}}, d_{\mathcal{X}}), (\alpha_{\mathcal{Y}}, d_{\mathcal{Y}}))^2 := \min_{\pi \in \mathcal{U}(\alpha_{\mathcal{X}}, \alpha_{\mathcal{Y}})} \int_{\mathcal{X}^2 \times \mathcal{Y}^2} |d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y')|^2 d\pi(x, y) d\pi(x', y')$$

es una distancia entre espacios métricos con medida salvo isometrías, donde $(\mathcal{X}, d_{\mathcal{X}}, \alpha_{\mathcal{X}})$ es isométrico a $(\mathcal{Y}, d_{\mathcal{Y}}, \alpha_{\mathcal{Y}})$ si existe una biyección $\phi : \mathcal{X} \rightarrow \mathcal{Y}$, tal que $\phi_{\#} \alpha_{\mathcal{X}} = \alpha_{\mathcal{Y}}$ y $d_{\mathcal{X}}(x, x') = d_{\mathcal{Y}}(\phi(x), \phi(x'))$.



Capítulo 2

Inferencia de redes de fMRI

Los datos analizados en este trabajo fueron grafos sin pesos no dirigidos que representan la conectividad funcional entre regiones del cerebro. Estos grafos, también conocidos como *conectomas*, son construidos a partir de imágenes fMRI con un método de inferencia desarrollado en [4]. Los conectomas analizados en esta tesis fueron compartidos por la autora de dicho artículo y colaboradora de este trabajo, la Dra. Sophie Achard. En las siguientes secciones describiremos las características y el método de construcción de los conectomas a partir de imágenes fMRI.

2.1 Descripción de los datos

Fueron dos conjuntos de datos analizados.

El primer conjunto de datos consiste en conectomas recolectados de treina voluntarios, categorizados en dos grupos dependiendo de su edad: el primer grupo fueron personas de 18 a 33 años de edad, mientras que el segundo consistió de personas de 62 a 76 años. En total fueron 17 jóvenes (9 hombres y 8 mujeres) y 13 adultos (6 hombres y 7 mujeres) [15].

El segundo conjunto de datos son conectomas de 25 pacientes en coma de un rango de edad de 21-82 años, y conectomas de 20 personas sanas, en el rango de 25 a 51 años de edad [2].

Los conectomas de ambos conjuntos de datos son grafos sin pesos no dirigidos con 90 nodos. En la próxima sección se describe el proceso de inferencia de dichos conectomas a partir de imágenes fMRI.

2.2 Inferencia de conectomas de fMRI

Cada participante fue escaneado acostado en reposo con los ojos cerrados durante un periodo determinado de tiempo (*entre 10 y 20 minutos*). Se adquirieron las imágenes en eco del gradiente con lectura ecoplanar (EPI) que representan el contraste dependiente del nivel de oxígeno en la sangre.

Después de obtener los datos de EPI, se corrigieron los desplazamientos geométricos causados por el movimiento de la cabeza durante el escaneo, utilizando la plantilla

de EPI del Instituto Neurológico de Montreal. Después, se hace una parcelación por regiones usando la plantilla anatómica de [16]. Esta parcelación divide cada hemisferio del cerebro en 45 partes anatómicas. Después se promedian las series de tiempo fMRI sobre todos los vóxeles de cada una de las regiones anatómicas, obteniendo 90 series de tiempo regionales. Cada serie de tiempo regional es corregida de nuevo por los efectos del movimiento de la cabeza. Para dicha corrección se aplicó una regresión sobre las series de tiempos de la traslación y rotación de la cabeza, las que fueron estimadas durante la primer corrección geométrica con realineamiento de imagen. Los residuos de estas regresiones constituyen el conjunto de series de tiempo promedio usados en el análisis de correlación con **transformadas wavelet**, definidas a continuación.

Transformada Wavelet

Definición 71. Una **función wavelet madre** es una función compleja ψ que satisface

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty$$

y

$$c_\psi = 2\pi \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$$

donde Ψ es la transformada de Fourier de ψ .

Definición 72. Sea $s : \mathbb{R} \rightarrow \mathbb{R}$ una función. Si ψ es una función wavelet madre, entonces la **transformada wavelet** de $s(t)$ con respecto a $\psi(t)$ está definida como

$$S(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi' \left(\frac{t-b}{a} \right) s(t) dt$$

donde $a > 0$, $b \in \mathbb{R}$ y ψ' es la conjugada compleja de ψ .

Las transformadas wavelet efectúan una descomposición en la escala de tiempo, partiendo la energía total de una señal sobre el conjunto de funciones wavelet madre, donde cada función es escalada en frecuencia y ubicada en el tiempo de forma única. Estas transformadas son particularmente adecuadas para señales de fMRI de personas en reposo, por las propiedades que conllevan dichas señales. Para mayor información sobre transformadas wavelet en datos fMRI véase [17], y para conocer la teoría general de análisis wavelet véase [18].

La transformada wavelet es la base que se usa para estimar las correlaciones entre las 90 series de tiempo regionales de fMRI, obteniendo una matriz de correlación inter-regional por wavelet de tamaño 90×90 .

A partir de esta matriz de correlación se construye un grafo no dirigido, cuyos nodos representan las 90 regiones anatómicas y están conectados por una arista si la correlación wavelet entre la región i y la región j excede un umbral predefinido.

Este método de construcción de redes a partir de imágenes de fMRI ha sido utilizado en una gran cantidad de estudios de redes neurológicas, superando 2,000 citas. En particular, en [15] utilizaron esta técnica para analizar las redes neuronales de personas en coma y personas control, sin embargo, en la sección de discusión dicen que no han



encontrado evidencia aparente de distinción estructural entre conectomas de estos dos grupos de personas. Encontrar la solución de este problema es la motivación de esta tesis. Aquí, desarrollamos una distancia que sí logre diferenciar entre los conectomas.



Capítulo 3

Resultados

La distancia espectral, la distancia de Wasserstein y la distancia de Gromov-Wasserstein definen los algoritmos para calcular la distancia entre conectomas. A continuación, describiremos estos algoritmos con cada una de las distancias mencionadas.

3.1 Distancia espectral truncada

La distancia espectral entre conectomas consiste en primero encontrar el vector de autovalores de la matriz sin retroceso de cada conectoma. Ya que la cantidad de autovalores no repetidos puede variar dependiendo del conectoma analizado, los vectores se truncan de la manera que ambos queden en el mismo espacio Euclideo. Después se calcula la distancia Euclidea entre los vectores truncados. Para remediar el hecho de que, por el teorema 34 de Constantine-Lafont, la longitud espectral ignora los nodos de grado uno de un grafo, agregamos un nodo cono. Considerando esto, calculamos la distancia espectral truncada desde tres perspectivas:

- Conectomas en su estado original
- Conectomas con cono conectado a los nodos de grado 1
- Conectomas con cono conectado a todos los nodos

El pseudocódigo de la distancia espectral truncada es el siguiente:

Algorithm 3.1.1 `spectral_distance(graphs, cone, method)`

Input: Lista de grafos, una variable booleana y el método de adición, que puede tomar los valores ‘all’ y ‘degree_one’

Output: Matriz de distancias

```

1: procedure
2:   graphs  $\leftarrow$  lista de conectomas
3:   m  $\leftarrow$  cantidad de conectomas por comparar
4:   cone  $\leftarrow$  variable booleana para determinar si se agrega un cono a los grafos
5:   method  $\leftarrow$  regla de adición del cono
6:   distancias  $\leftarrow$  matriz  $m \times m$  con ceros
7:   if cone then
8:     for  $0 \leq j < m$  do
9:       graphs[j]  $\leftarrow$  add_cone(graphs[j], method) ▷ Ver algoritmo 1.3.1
10:    end for
11:  end if
12:  eigs  $\leftarrow$  nbeigs_calculate(graphs) ▷ Ver algoritmo 1.3.2
13:  for  $0 \leq i \leq j < m$  do ▷ Calcular distancia entre cada pareja de grafos
14:    n  $\leftarrow$  mínimo de las longitudes de eigs[i] y eigs[j]
15:    distancias[i, j]  $\leftarrow$  distancia euclideana entre eigs[i][:n] y eigs[j][:n]
16:  end for
17:  return distancias
18: end procedure

```

Utilizando este algoritmo, graficamos los mapas de calor de las distancias resultantes. Se obtienen tres mapas de calor por cada conjunto de datos analizados (véase Figura 3.1).

Se observa que los mapas de calor representando las distancias entre conectomas con cono tienen una escala más amplia, distinguiendo de mejor manera a los conectomas entre sí. Cabe notar que la regla de conexión del cono con todos los nodos no hace mucha mejora, mientras que la conexión del cono con nodos de grado uno hace la distinción entre conectomas más notoria.

La desventaja de utilizar la distancia espectral es el hecho de tener que truncar los vectores característicos para convertirlos en vectores del mismo espacio, lo que conlleva pérdida de información del conectoma.

3.2 Wasserstein KDE

Una forma de evitar el truncamiento de vectores y la pérdida de información es estimar la función de densidad de los autovalores de la matriz sin retrocesos de cada conectoma. Así, obtenemos funciones que, por construcción, están en el mismo espacio. La distancia de Wasserstein permite comparar dichas funciones.

Ya que los cálculos a computadora están limitados a representaciones finitas de funciones, es necesario discretizar la función de densidad. Hay dos formas de hacerlo:



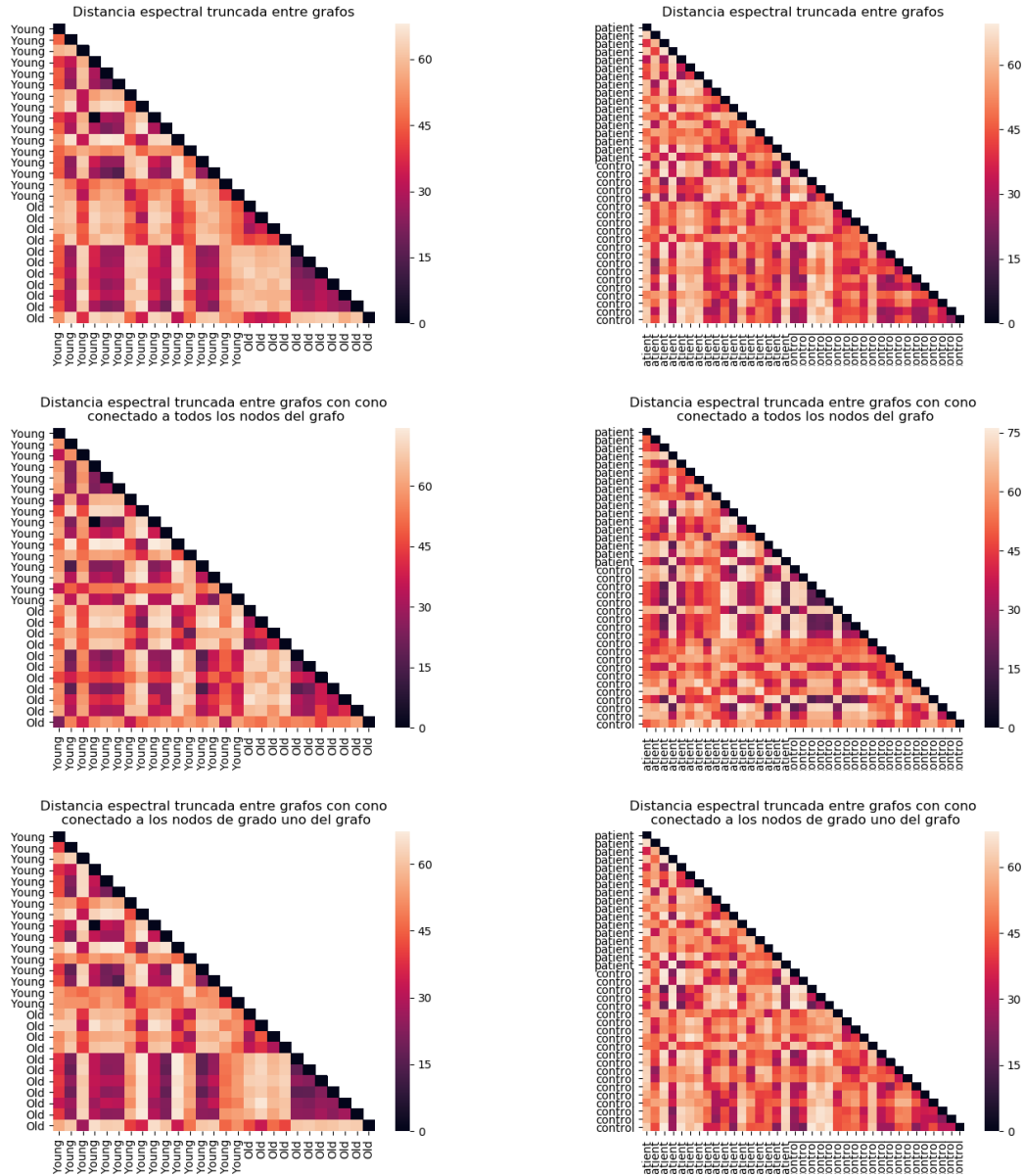


Figura 3.1: Distancia espectral entre pacientes en coma y personas control. La columna izquierda compara personas jóvenes y mayores, y la derecha son pacientes en coma y personas control.



de manera muestral o partiendo su dominio en una malla. Describiremos cada una de estas formas a continuación.

3.2.1 Wasserstein KDE muestral

Sean G y H dos conectomas. Denotemos por f_G y f_H a la estimación por núcleo de la función de densidad de los autovalores de la matriz sin retroceso del conectoma G y el conectoma H , respectivamente. Para calcular la distancia entre f_G y f_H , se extraen dos muestras de tamaño n con distribución f_G y f_H . Después, se calcula la distancia de Wasserstein entre ambas muestras.

El pseudocódigo es el siguiente:

Algorithm 3.2.1 `wasserstein_kde_distance(graphs, cone, ker, bw, sample_size)`

Input: Lista de grafos y una variable booleana

Output: Matriz de distancias

```

1: procedure
2:   graphs  $\leftarrow$  lista de conectomas
3:    $m \leftarrow$  cantidad de conectomas por comparar
4:   cone  $\leftarrow$  Variable booleana para determinar si se agregan conos a los grafos
5:   ker  $\leftarrow$  núcleo para la estimación de la función de densidad con núcleo
6:   bw  $\leftarrow$  ancho de ventana para la estimación de la función de densidad con núcleo
7:   sample_size  $\leftarrow$  tamaño de muestra por extraer
8:   distancias  $\leftarrow$  matriz  $m \times m$  con ceros
9:   if cone then
10:    for  $0 \leq j < m$  do
11:      graphs[j]  $\leftarrow$  add_cone(graphs[j]) ▷ Ver algoritmo 1.3.1
12:    end for
13:  end if
14:  eigs  $\leftarrow$  nbeigs_calculate(graphs) ▷ Ver algoritmo 1.3.2
15:  models  $\leftarrow$  lista vacía de tamaño  $m$ 
16:  for  $0 \leq j < m$  do ▷ Estimamos los modelos
17:    models[j]  $\leftarrow$  estimación de la función de densidad de eigs[j] con núcleo ker
    y ancho de ventana bw
18:  end for
19:  samples  $\leftarrow$  lista vacía de tamaño  $m$ 
20:  for  $j < m$  do ▷ Sacamos la muestra dada la función de densidad estimada
21:    samples[j]  $\leftarrow$  muestra  $[x_1, \dots, x_{sample\_size}]$  de distribución models[j]
22:  end for
23:  for  $0 \leq i < j < m$  do ▷ Calcular distancia entre cada pareja de grafos
24:     $C \leftarrow$  matriz de distancia normalizada entre la muestra samples[i] y samples[j]
25:     $p \leftarrow$  histograma uniforme de tamaño sample_size
26:     $q \leftarrow$  histograma uniforme de tamaño sample_size

```



```
27:     distancias[i, j] ← distancia de Wasserstein entre p y q dada la función
      de costo C
28:   end for
29:   return distancias
30: end procedure
```

Los resultados de aplicar Wasserstein KDE muestral para calcular la distancia entre conectomas se muestra en la Figura 3.2. Al igual que en la sección anterior, son tres mapas de calor por cada conjunto de datos. Vemos que las distancias entre conectomas son muy parecidas entre sí. Para el conjunto de conectomas de personas jóvenes y mayores no hace mucha diferencia la adición de un cono, pues vemos que la escala es muy similar entre ellas. En el caso de las distancias entre conectomas de pacientes en coma y personas control, la adición de un nodo disminuye la distinción entre conectomas. El mapa de calor de conectomas con cono conectado a todos los nodos del grafo es el que muestra las distancias más chicas entre conectomas, implicando que la adición de un cono conectado a todos los nodos disminuye las diferencias topológicas entre los conectomas.



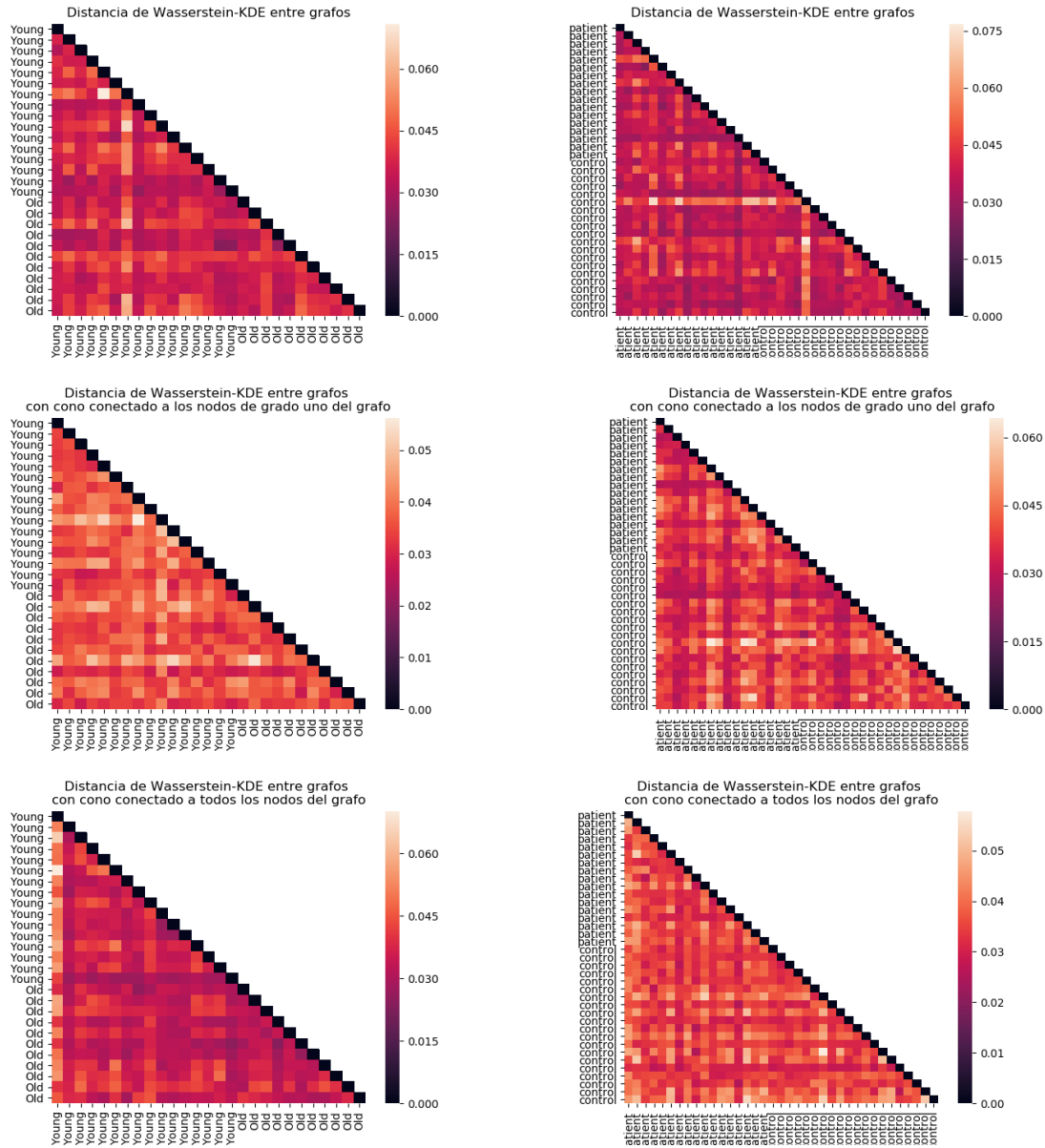


Figura 3.2: Distancia de Wasserstein KDE vía muestreo. La columna izquierda compara personas jóvenes y mayores, y la de la derecha es de pacientes en coma y personas control.

3.2.2 Wasserstein KDE en malla

Sean G y H dos conectomas. Denotemos por $f_G, f_H : \mathbb{C} \rightarrow \mathbb{R}$ a las estimaciones de la función de densidad de los autovalores de la matriz sin retroceso de los conectomas G y H , respectivamente. Para discretizar las funciones de densidad, dividamos el dominio en una malla regular. Las aproximaciones que tomaremos serán los valores de f_G y f_H en las intersecciones de la malla. Entre más densa es la malla, la aproximación a la función de densidad será más acertada, pero la complejidad del algoritmo aumenta.



El pseudocódigo para Wasserstein KDE en malla es el siguiente:

Algorithm 3.2.2 `wasserstein_kde_distance(graphs, cone, ker, bw, grid_size)`

Input: Lista de grafos y una variable booleana

Output: Matriz de distancias

```

1: procedure
2:   graphs  $\leftarrow$  lista de conectomas
3:    $m \leftarrow$  cantidad de conectomas por comparar
4:   cone  $\leftarrow$  Variable booleana para determinar si se agregan conos a los grafos
5:   ker  $\leftarrow$  núcleo para la estimación de la función de densidad con núcleo
6:   bw  $\leftarrow$  ancho de ventana para la estimación de la función de densidad con núcleo
7:   grid_size  $\leftarrow$  tamaño de muestra por extraer
8:   distancias  $\leftarrow$  matriz  $m \times m$  con ceros
9:   if cone then
10:     for  $0 \leq j < m$  do
11:       graphs[j]  $\leftarrow$  add_cone(graphs[j]) ▷ Ver algoritmo 1.3.1
12:     end for
13:   end if
14:   eigs  $\leftarrow$  nbeigs_calculate(graphs) ▷ Ver algoritmo 1.3.1
15:   models  $\leftarrow$  lista vacía de tamaño  $m$ 
16:   for  $0 \leq j < m$  do ▷ Estimamos los modelos
17:     models[j]  $\leftarrow$  estimación de la función de densidad de eigs[j] con núcleo ker
    y ancho de ventana bw
18:   end for
19:   grids  $\leftarrow$  lista vacía de tamaño  $m$ 
20:   for  $0 \leq j < m$  do ▷ Construimos la malla para discretizar la función de distribución
21:     grid[j]  $\leftarrow$  malla del dominio de models[j]
22:   end for
23:   for  $0 \leq i < j < m$  do ▷ Calcular distancia entre cada pareja de grafos
24:      $vector_i \leftarrow$  models[i] evaluado en cada uno de los puntos de grid
25:      $vector_j \leftarrow$  models[j] evaluado en cada uno de los puntos de grid
26:      $C \leftarrow$  matriz de distancia normalizada entre  $vector_i$  y  $vector_j$ 
27:      $p \leftarrow$  histograma uniforme de tamaño grid_size
28:      $q \leftarrow$  histograma uniforme de tamaño grid_size
29:     distancias[i, j]  $\leftarrow$  distancia de Wasserstein entre  $p$  y  $q$  dada la función
    de costo  $C$ 
30:   end for
31:   return distances
32: end procedure

```

Los resultados del algoritmo se muestran en la Figura 3.3.

A diferencia de Wasserstein-KDE muestral, la adición de un cono al conectoma ayuda a la discriminación de conectomas. A saber, la escala de los mapas de calor es



más amplia cuando se agregan conos a los grafos.

Obsérvese que para el conjunto de personas jóvenes y viejas, la distancia de Wasserstein KDE aplicada logra discriminar la categoría de un conjunto de conectomas. En efecto, los valores de la distancia son, en general, muy bajos entre conectomas de la misma categoría, y más altos entre conectomas de diferentes categorías. Este fenómeno no es tan evidente para la base de datos de personas en coma y control, pero la adición del cono consistentemente ayuda a la discriminación entre conectomas.

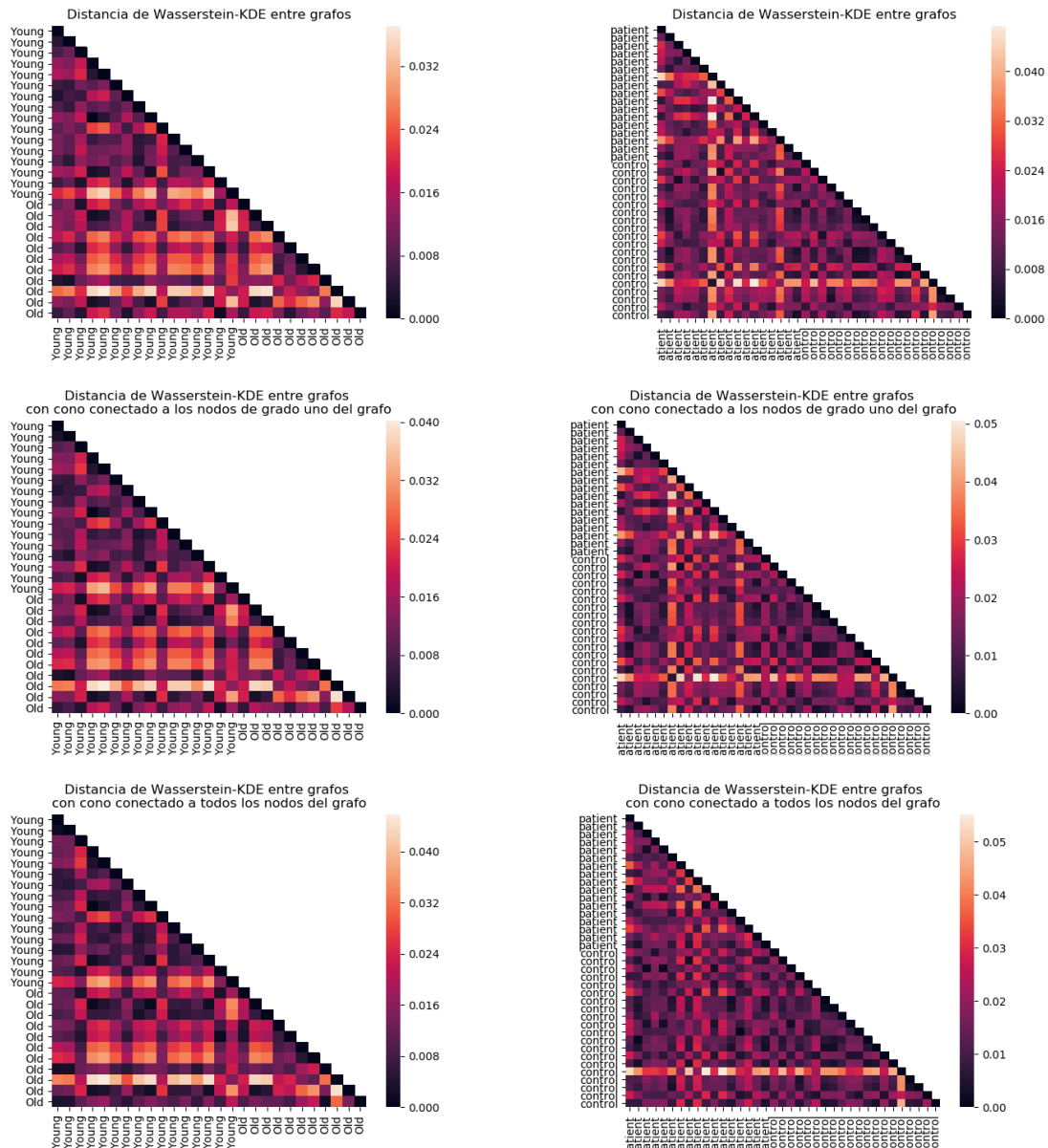


Figura 3.3: Distancia de Wasserstein KDE vía malla. La columna izquierda compara personas jóvenes y mayores, y la de la derecha es de pacientes en coma y personas control.



3.3 Gromov-Wasserstein

Por último, sean G y H dos conectomas entre los cuales queremos calcular la distancia. Sean $a = (a_1, \dots, a_n)$ y $b = (b_1, \dots, b_m)$ los vectores de autovalores de las matrices sin retrocesos de G y H , respectivamente. Denotemos por $M_a \in \mathbb{R}^{n \times n}$ a la matriz de distancia entre las entradas de a , y $M_b \in \mathbb{R}^{m \times m}$ la matriz de distancia entre las entradas de b . La distancia de Gromov-Wasserstein entre los conectomas G y H es entonces $GW((a, M_a), (b, M_b))$, definida en la sección 1.5.2

El pseudocódigo de este algoritmo es el siguiente:

Algorithm 3.3.1 distance_gr_wass(*graphs, cone*)

Input: Lista de grafos y una variable booleana

Output: Matriz de distancias

```

1: procedure
2:   graphs ← lista de conectomas
3:   m ← cantidad de conectomas por comparar
4:   cone ← Variable booleana para determinar si se agregan conos a los grafos
5:   distancias ← matriz m × m con ceros
6:   if cone then
7:     for 0 ≤ j < m do
8:       graphs[j] ← add_cone(graphs[j])           ▷ Ver algoritmo 1.3.1
9:     end for
10:  end if
11:  eigs ← nbeigs_calculate(graphs)                ▷ Ver algoritmo 1.3.2
12:  for 0 ≤ i ≤ j < m do                            ▷ Calcular distancia entre cada pareja de grafos
13:    C1 ← matriz de distancias entre cada entrada de eigs[i] normalizada
14:    C2 ← matriz de distancias entre cada entrada de eigs[j] normalizada
15:    n ← tamaño del vector eigs[i]
16:    p ← histograma uniforme de tamaño n
17:    m ← tamaño del vector eigs[j]
18:    q ← histograma uniforme de tamaño n
19:    distancias[i,j] ← distancia de Gromov-Wasserstein entre p con matriz de
    similaridad C1 y q con matriz de similaridad C2
20:  end for
21:  return distancias
22: end procedure

```



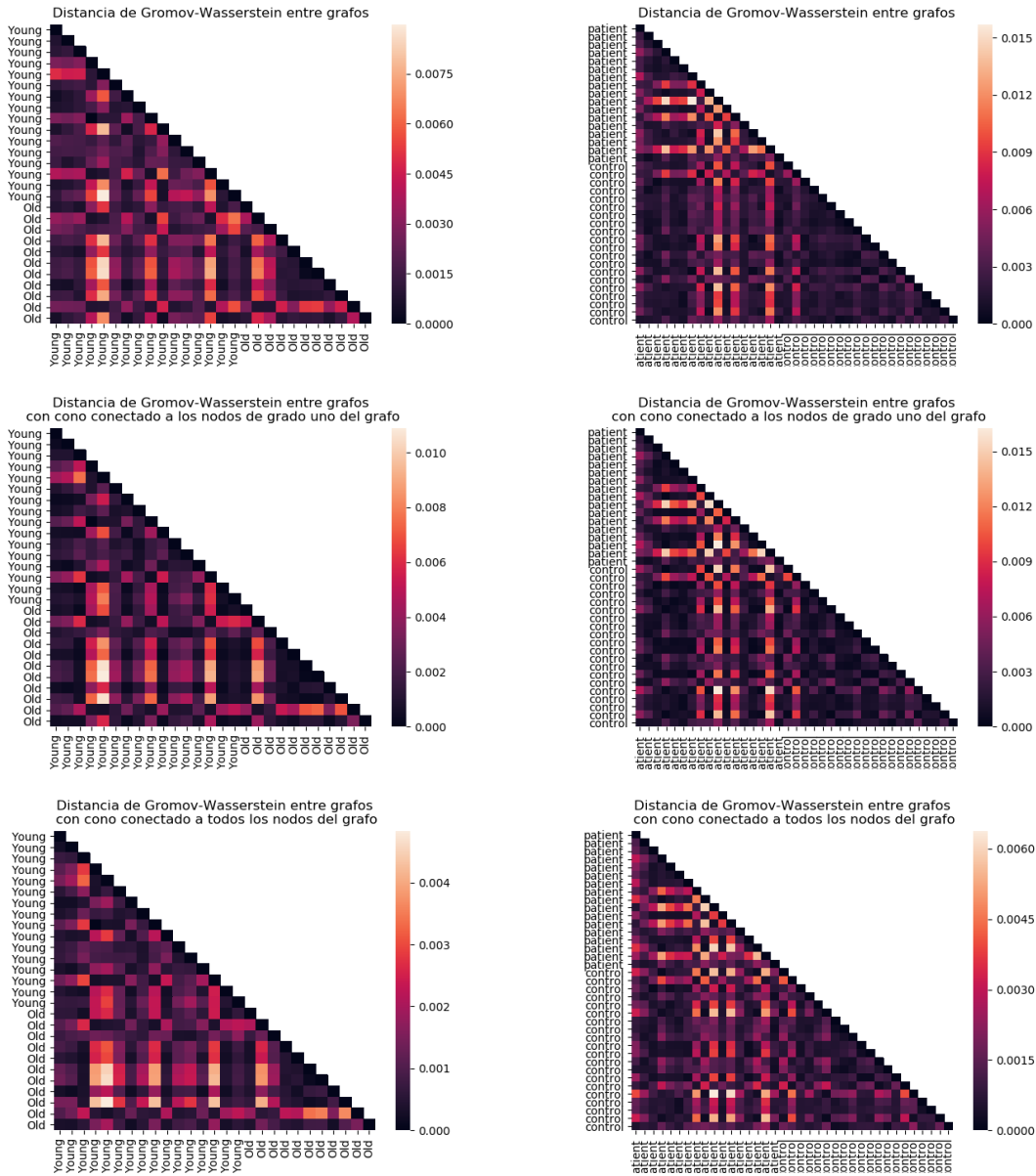


Figura 3.4: Distancia de Gromov-Wasserstein entre conectomas. La columna izquierda es la comparación de personas jóvenes y mayores, y la de la derecha es de pacientes en coma y personas control

Utilizando este algoritmo, la distancia entre conectomas se aprecia en la Figura 3.4. Aquí vemos que el algoritmo de Gromov-Wasserstein es el que menos logra discriminar entre conectomas. Sin embargo, la consecuencia de agregar un cono es más marcada. Cuando se agrega un cono conectado a los nodos de grado uno al grafo, la discriminación entre conectomas aumenta por un orden de magnitud para el conjunto de personas jóvenes y viejas, mientras que si el cono está conectado a todos los nodos, la discriminación disminuye. La tendencia se repite en los mapas de calor entre conectomas de pacientes en coma y personas control.



Concluimos que la regla de conexión del cono al grafo es importante, pues cuando el cono se conecta solo a los nodos de grado 1, se logra discriminar de mejor manera los conectomas, ya que las escalas de los mapas de calor son más amplias, mientras que el cono conectado a todos los nodos hace que el grafo sea menos distinguible de los demás.



Capítulo 4

Conclusiones

Tres algoritmos fueron propuestos para cuantificar la diferencia entre conectomas: la distancia espectral truncada, la distancia de Wasserstein KDE y la distancia de Gromov Wasserstein. Cada uno de los algoritmos fue analizado desde tres perspectivas: aplicándolos sobre conectomas originales y sobre conectomas con cono. Se siguieron dos reglas de conexión del cono, conectándolo a todos los nodos, o solamente a los nodos de grado uno.

La distancia espectral entre conectomas aumenta cuando se le agrega un cono siguiendo la regla mínima, y se mantiene igual cuando la regla de conexión es máxima.

La distancia de Wasserstein KDE muestral fue la menos esclarecedora, pues todas las distancias eran muy parecidas y mostraban un comportamiento casi uniforme. La adición del cono no cambió la situación.

La distancia de Wasserstein-KDE vía malla logra diferenciar los conectomas de personas jóvenes y viejas. Este fenómeno no es tan evidente para la base de datos de personas en coma y control. La adición del cono remarca la diferencia entre conectomas para ambos conjuntos, haciendo las distancias más grandes.

Por último, la distancia de Gromov-Wasserstein demuestra la importancia de la regla de conexión elegida al añadir un cono al grafo. Cuando el cono es conectado a los nodos de grado 1, la distancia entre conectomas es más alta. De lo contrario, cuando la conexión del cono es a todos los nodos del grafo, los grafos son parecidos entre sí desde el punto de vista topológico, ya que la distancia entre ellos disminuye.

En la mayoría de los algoritmos, notamos que agregar un cono a los grafos logra discriminar de mejor manera los conectomas. En efecto, las escalas de los mapas de calor son más amplias cuando se agregan los conos. Sin embargo, la regla de adición juega un rol importante, pues cuando el cono se conecta a los nodos de grado uno, hace la diferencia entre conectomas más marcada, mientras que a todos los nodos los hace más parecidos entre sí.

Obsérvese que la distancia de Wasserstein KDE aplicada sobre la base de datos de personas jóvenes y mayores logra discriminar la categoría del conectoma. En efecto, los valores de la distancia son muy bajos entre conectomas de la misma categoría, y más altos entre conectomas de diferentes categorías.

Este trabajo abre varias posibilidades de seguir la investigación, como el análisis de bases de datos con más categorías, cubriendo un mayor espectro de enfermedades

neurológicas, o aplicar estas técnicas en problemas de redes en otros contextos.

Otra línea interesante de analizar es la elección de conexiones que se hacen del nodo cono a los nodos del grafo. Hemos visto que con diferentes reglas de conexión, la distancia entre redes puede ser más marcada, o al revés. Por lo tanto, la regla de conexión del cono con los demás nodos da diferentes resultados, y esto es una oportunidad para mejorar la diferenciación entre categorías de redes.



Bibliografía

- [1] A. J. Solomon, R. T. Naismith, and A. H. Cross, “Misdiagnosis of multiple sclerosis,” *Neurology*, vol. 92, no. 1, pp. 26–33, 2019. [iii](#)
- [2] S. Achard, C. Delon-Martin, P. E. Vértes, F. Renard, M. Schenck, F. Schneider, C. Heinrich, S. Kremer, and E. T. Bullmore, “Hubs of brain functional networks are radically reorganized in comatose patients,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 50, pp. 20608–20613, 2012. [iii](#), [iv](#), [25](#)
- [3] L. Torres, P. Suárez-Serrato, and T. Eliassi-Rad, “Non-backtracking cycles: length spectrum theory and graph mining applications,” *Applied Network Science*, vol. 4, no. 41, 2019. [iii](#), [11](#)
- [4] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, “A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs,” *J Neurosci*, vol. 3, p. e17, 01 2006. [iv](#), [25](#)
- [5] A. Hatcher, C. U. Press, and C. U. D. of Mathematics, *Algebraic Topology*. Algebraic Topology, Cambridge University Press, 2002. [5](#)
- [6] D. Constantine and J.-F. Lafont, “Marked length rigidity for one dimensional spaces,” *Journal of Topology and Analysis*, 09 2012. [7](#)
- [7] S. Lang, *Linear Algebra*. Springer Undergraduate Texts in Mathematics and Technology, Springer, 1987. [12](#)
- [8] A. Mood, A. Mood, D. BOES, F. Graybill, D. Boes, D. Boes, and G. A, *Introduction to the Theory of Statistics*. International Student edition, McGraw-Hill, 1973. [14](#)
- [9] Y.-C. Chen, “A tutorial on kernel density estimation and recent advances,” *Biostatistics Epidemiology*, vol. 1, 04 2017. [16](#)
- [10] C. Villani, *Optimal transport – Old and new*, vol. 338, pp. xxii+973. 01 2008. [16](#)
- [11] G. Peyré and M. Cuturi, *Computational Optimal Transport: With Applications to Data Science*. 2019. [16](#), [21](#), [24](#)
- [12] G. Monge, “Mémoire sur la théorie des déblais et des remblais,” *Histoire de l’Académie Royale des Sciences de Paris*, pp. 666–704, 1781. [16](#)

- [13] L. Kantorovich, “On the transfer of masses (in russian),” *Doklady Akademii Nauk*, vol. 37, p. 227–229, 1942. [20](#)
- [14] F. Mémoli, “Gromov–wasserstein distances and the metric approach to object matching,” *Foundations of Computational Mathematics*, vol. 11, pp. 417–487, 08 2011. [23](#)
- [15] S. Achard and E. Bullmore, “Efficiency and cost of economical brain functional networks,” *PLoS Comput Biol*, vol. 3, 2007. [25](#), [26](#)
- [16] N. Tzourio-Mazoyer, B. Landeau, P. DF, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and J. Marc, “Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain,” *NeuroImage*, vol. 15, pp. 273–89, 02 2002. [26](#)
- [17] E. Bullmore, J. Fadili, V. Maxim, L. Sendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear, “Wavelets and functional magnetic resonance imaging of the human brain,” *NeuroImage*, vol. 23 Suppl 1, pp. S234–49, 02 2004. [26](#)
- [18] D. Percival, A. Walden, R. Gill, B. Ripley, S. Ross, D. Williams, M. Stein, and B. Silverman, *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2000. [26](#)

