
Interacciones de orden superior: hipergrafos y complejos simpliciales

PID_00286324

Lluís Arola Fernández

Tiempo mínimo de dedicación recomendado: 2 horas



**Lluís Arola Fernández**

Graduado en Física, con mención en Física Fundamental, por la Universidad de Barcelona (UB) y en el máster Modelling for Science and Engineering, con especialidad en Sistemas Complejos, de la Universidad Autónoma de Barcelona (UAB). Después de trabajar en proyectos industriales en el Institut de Recerca d'Energia de Catalunya (IREC), en 2022 obtuvo el doctorado en la Universidad Rovira i Virgili (URV) bajo la supervisión del Dr. Àlex Arenas, con una tesis centrada en la teoría de la sincronización en redes complejas. Actualmente es investigador posdoctoral en el Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), centro del CISC adscrito a la Universitat de les Illes Balears (UIB), donde desarrolla modelos matemáticos en la intersección entre la física teórica y el *deep learning*.

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por el profesor: Jordi Casas Roma

Cómo citar este recurso de aprendizaje con el estilo Harvard:

Arola Fernández, L. (2023) *Interacciones de orden superior: hipérgrafos y complejos simpliciales*. [Recurso de aprendizaje textual]. 1.^a ed. Barcelona: Fundació Universitat Oberta de Catalunya (FUOC).

Primera edición: septiembre 2023
© de esta edición, Fundació Universitat Oberta de Catalunya (FUOC)
Av. Tibidabo, 39-43, 08035 Barcelona
Autoria: Lluís Arola Fernández
Producción: FUOC
Todos los derechos reservados

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea este eléctrico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita del titular de los derechos.

Índice

Introducción	5
Objetivos	6
1. Conceptos preliminares.....	7
2. Marco teórico	8
2.1. Definiciones: hipergrafos y complejos simpliciales	8
2.2. Representaciones matriciales	10
2.3. Métricas de centralidad	13
3. Aplicaciones	15
3.1. Estructura: modelos geométricos y análisis topológico de datos	15
3.2. Dinámica: efectos abruptos en transiciones de fase	18
Resumen	22
Glosario	23
Bibliografía	24

Introducción

En este material didáctico introduciremos los conceptos básicos para entender los objetos conocidos como hipergrafos y complejos simpliciales (en inglés, *hypergraphs* y *simplicial complexes*), que generalizan las ideas de los grafos a objetos que permiten codificar y analizar interacciones más allá de por pares, es decir, entre tres o más nodos simultáneamente. Este campo de estudio de grafos con interacciones de orden superior se ha beneficiado de mucha investigación realizada en matemáticas puras y aplicadas (por ejemplo, en topología algebraica y en el análisis topológico de conjuntos de datos), en la física de los sistemas complejos y la teoría de redes, y en campos más aplicados como la sociología o la epidemiología.

En el apartado 1 de este material, discutiremos las limitaciones de los grafos simples e introduciremos las motivaciones y ventajas principales de estudiar redes con interacciones de orden superior en contextos sociales y en análisis de datos relacionales.

A continuación, en el apartado 2, expondremos el marco teórico necesario para entender estos objetos y estudiarlos matemáticamente. En particular, en el subapartado 2.1 definiremos los hipergrafos de forma general y hablaremos de una clase de hipergrafos que recibe un interés especial, los complejos simpliciales. En el subapartado 2.2, veremos algunas de las principales representaciones matriciales y tensoriales que describen estos objetos y que permiten extender el concepto de matriz de adyacencia a órdenes superiores. En el subapartado 2.3, usaremos estas representaciones para calcular algunas de las métricas de centralidad de los nodos en presencia de interacciones de orden superior.

Seguidamente, en el apartado 3, introduciremos algunas de las aplicaciones más importantes de los hipergrafos y los complejos simpliciales. En el subapartado 3.1, veremos cómo los modelos geométricos de hipergrafos y el análisis de sus propiedades topológicas a distintas escalas de resolución permiten capturar la geometría de un conjunto de datos de una manera mucho más precisa y sutil que usando grafos simples. Después de discutir las aplicaciones basadas en la estructura, en el subapartado 3.2 veremos cómo se pueden introducir las interacciones de orden superior en el modelaje de procesos dinámicos, por ejemplo, en un contagio epidémico o un proceso de sincronización. Descubriremos que las interacciones de orden superior pueden facilitar la aparición de transiciones abruptas en variables macroscópicas del sistema, por ejemplo, en la prevalencia de una epidemia o una idea o en el grado de sincronización global.

Objetivos

En este material didáctico encontraremos las herramientas indispensables para asimilar los siguientes objetivos:

- 1.** Entender las limitaciones de las redes y los grafos clásicos para capturar relaciones e interacciones existentes en sistemas reales y las motivaciones para el uso de redes o grafos con interacciones de orden superior.
- 2.** Comprender las definiciones principales y las distintas clases de redes con interacciones de orden superior, incluyendo los hipergrafos y los complejos simpliciales como una clase particular de estos.
- 3.** Conocer las principales representaciones matriciales y tensoriales de estas redes y algunas de las extensiones de métricas y modelos de redes a estos grafos más generales.
- 4.** Entender el importante uso de los hipergrafos y complejos simpliciales en el contexto del análisis topológico de datos. En particular, veremos como los modelos geométricos generativos de hipergrafos y la filtración de complejos simpliciales a distintas escalas de resolución permiten aplicar análisis de homología persistente y también discutiremos como los modelos nulos probabilísticos pueden ser útiles para explicar datos relacionales en sistemas sociales.
- 5.** Descubrir como la inclusión de redes con interacciones de orden superior en el modelaje de procesos dinámicos puede afectar drásticamente los fenómenos dinámicos del modelo. En particular, veremos como la presencia de interacciones de orden superior (más allá de por pares) puede inducir transiciones abruptas (explosivas) tanto en procesos de sincronización como en dinámicas de contagio «complejo», ya sea epidémico o social.

1. Conceptos preliminares

A lo largo de los materiales didácticos de este curso, hemos comprendido la importancia y la utilidad que tienen los grafos para modelar muchos sistemas reales, como por ejemplo las redes sociales. Hemos visto las métricas y medidas principales de los grafos, y como estos nos ayudan a analizar procesos dinámicos que ocurren en los vértices, como la propagación de información y a abordar problemas cruciales como el análisis de comunidades. También hemos extendido los grafos más simples en muchas direcciones para capturar efectos más sutiles y realistas más allá de interacciones estáticas, no dirigidas y en monocapa, por ejemplo, en el contexto de las redes multicapa o las redes temporales.

Una limitación fundamental de los grafos, ya sea monocapa o multicapa, estáticos o temporales, es que estos solo capturan relaciones entre parejas de vértices. No obstante, en muchos sistemas sociales, ecológicos o biológicos, entre muchos otros ejemplos, hay muchas interacciones y relaciones que no suceden entre parejas de nodos, sino entre grupos de vértices, como acciones colectivas. Por ejemplo, tres o más especies pueden competir por comida y la dominación de un territorio en un ecosistema complejo, o la presencia de una tercera especie puede influir en la interacción entre otras dos. En otro contexto más social, mecanismos como por ejemplo la presión social (*peer-pressure*) o el contagio complejo solo se pueden explicar mediante relaciones que van más allá de interacciones por pares. Finalmente, en un contexto más aplicado y basado en el análisis de datos, la construcción de grafos con interacciones a pares entre los datos puede no ser suficiente para capturar la geometría y las interacciones de los datos de una forma más robusta e interpretable.

¿Cómo podemos definir y representar matemáticamente estos grafos más generales? ¿Podemos extender las distintas métricas de centralidad de los nodos en presencia de interacciones de orden superior? ¿Cuáles son las aplicaciones e implicaciones más importantes de estos objetos en el contexto de las redes sociales o en los procesos de propagación de información? Estas son algunas de las preguntas que trataremos de responder en este material didáctico. Es importante mencionar que expondremos ideas introductorias y tan solo algunos resultados fundamentales, sugiriendo referencias de la literatura que abordan los distintos temas expuestos en más profundidad. Como hemos dicho, el estudio de este tipo de objetos se ramifica en muchos campos de estudio, y en el contexto del análisis de grafos y redes sociales estamos aún al principio del camino; se trata de un tema en constante desarrollo en la actualidad. Quedan muchos problemas abiertos y preguntas interesantes por abordar en el futuro.

2. Marco teórico

2.1. Definiciones: hipergrafos y complejos simpliciales

Es conveniente empezar con las definiciones, que nos permitirán fijar la terminología que usaremos a lo largo del material. Para ello, en este apartado definiremos matemáticamente los objetos que capturan las interacciones de orden superior en grafos.

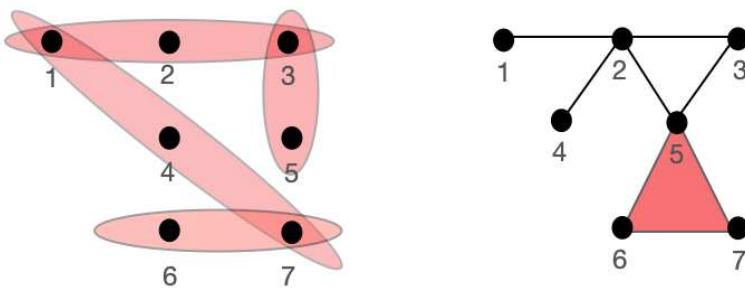
Antes de empezar, recordamos que un grafo simple no dirigido $G = (V, E)$ consiste, primero, en un set finito $V = \{v_1, \dots, v_n\}$ de vértices y, segundo, en un set $E = \{v_i, v_j\} : v_i, v_j \in V$ de aristas, que corresponden a parejas no ordenadas de elementos de V . En contraste, un grafo dirigido es un par $G = (V, E)$ de un set finito de vértices $V = \{v_1, \dots, v_n\}$ y un set de aristas $E \in V \times V$, que corresponde a pares ordenados de vértices. Cualquier grafo G puede ser identificado con una matriz de adyacencia $n \times n$, donde $A_{ij} = 1$ si $(i, j) \in E(G)$. Si el grafo es no dirigido, la matriz es simétrica, es decir $A = A^T$ (donde A^T es la transpuesta de la matriz). Esta idea se puede extender a grafos ponderados, donde las entradas de la matriz A capturan la intensidad de las interacciones, y a grafos dirigidos, donde $A \neq A^T$.

Habiendo recordado estos conceptos básicos de los grafos, extendemos estas ideas a unos grafos más generales, los hipergrafos. La figura 1 presenta dos ejemplos que definimos a continuación.

Figura 1. Ejemplos de hipergrafo y complejo simplicial

a) Hypergrafo

b) Simplicial complex



Ejemplos de hipergrafo y complejo simplicial

En la figura 1a podemos ver un ejemplo de hipergrafo con 7 vértices y 4 hiperaristas (*hyperedges*), mientras que en la figura 1b observamos un ejemplo de complejo simplicial con 7 vértices, 8 aristas (1-simplex) y un triángulo (2-simplex). En ambos casos, los grafos se asumen no dirigidos y no ponderados por simplicidad.

Un **hipergrafo** es una generalización del grafo no dirigido. Denominamos $P(X)$ el conjunto potencia del conjunto X , que consisten en todos los subconjuntos de X . Entonces, si $V = \{v_1, \dots, v_n\}$ es un conjunto finito de vértices y $E \in P(V)$ es un conjunto finito de subconjuntos no vacíos en V (lo que llamamos hiperaristas), el par $H = (V, E)$ es conocido como hipergrafo.

Como caso particular, un hipergrafo k -uniforme es aquel en el que todas las hiperaristas $e \in E$ tienen la misma cardinalidad $|e| = k$.

Ejemplo de hipergrafo

Consideramos $V = \{v_1, \dots, v_7\}$ como el conjunto de vértices y, por otro lado, el conjunto $E = \{\{1,2,3\}, \{1,4,7\}, \{3,5\}, \{6,7\}\}$, como el conjunto de hiperaristas. Este hipergrafo es ilustrado en la figura 1a. Por ejemplo, el vértice v_3 tiene una interacción de pareja con el vértice v_5 y una interacción a tres con los vértices v_1 y v_2 .

Es importante mencionar que los hipergrafos se pueden extender para incorporar direccionalidad y pesos, aunque no siempre de forma unívoca. En un contexto de grafos con interacciones de orden superior, los pesos (enteros) pueden representar las veces que una hiperarista aparece en un hipergrafo. También es interesante ver que en un hipergrafo general, una hiperarista de cardinalidad 3 (un triángulo) puede existir entre tres nodos sin que haya ninguna hiperarista de cardinalidad 2 (una arista) entre esos nodos. Una clase particular de hipergrafos que sí que presentan la restricción anterior son los complejos simpliciales, que definimos a continuación.

Un **complejo simplicial** es un hipergrafo donde el conjunto de hiperaristas es cerrado bajo inclusión, es decir, que cada subconjunto no vacío d de una hiperarista $e \in E$ es también un miembro del conjunto de hiperaristas.

En particular, si consideramos un conjunto finito de vértices $V = \{v_1, \dots, v_n\}$, definimos un m -simplex como $\sigma = \{v_0, \dots, v_m\}$, un conjunto no vacío de $m+1$ elementos de V y su dimensión viene dada por m . Dado un m -simplex σ , llamamos $\tau \in \sigma$ a la cara de σ . Entonces K es un complejo simplicial, es decir, una colección finita de los simplex donde $\tau \in \sigma$ y $\sigma \in K$ implica $\tau \in K$ y además $\{v\} \in K$ para todos los $v \in V$. La dimensión del complejo simplicial K se define como la dimensión mayor de cualquier simplex en K . El m -esqueleto K_m del complejo simplicial es la unión de los simplex de dimensiones $0, 1, \dots, m$. Por ejemplo, el 0-esqueleto $K_0 = V$, donde V son los vértices de K , y el 1-esqueleto es el grafo de interacciones a pares asociado con el complejo simplicial.

Es importante ver que, como los complejos simpliciales son un caso particular de hipergrafos, podemos usar la notación $K = (V, E) = (K_0, E)$, con V los elementos de K_0 y E siendo el conjunto de todos los simplices contenidos en K .

Ejemplo de complejo simplicial

Consideramos el conjunto de vértices $V = \{v_1, \dots, v_7\}$, y el conjunto de simplex $K = \{\{1\}, \dots, \{7\}, \{1,2\}, \{2,3\}, \{2,4\}, \{2,5\}, \{3,5\}, \{5,6\}, \{6,7\}, \{5,7\}, \{5,6,7\}\}$, ilustrado en la figura 1b. Este complejo simplicial K tiene dimensión 2 ya que $\{5,6,7\}$ es el simplex mayor en K .

2.2. Representaciones matriciales

De igual manera que ocurría con el caso de los grafos simples, es importante obtener representaciones matemáticas adecuadas de los hipergrafos. En los grafos o redes por parejas, hemos visto que la matriz de adyacencia es normalmente el objeto más adecuado y natural para describir las relaciones binarias. La matriz de adyacencia y algunas de sus matrices relacionadas, como el Laplaciano del grafo, permiten calcular la mayoría de métricas locales y globales de interés y también emergen de forma natural en el contexto de los modelos que describen procesos dinámicos en grafos.

En el caso de las redes con interacciones de orden superior, o simplemente hipergrafos, la elección de la representación no es tan directa y depende de la información disponible y de la aplicación en particular. En este subapartado veremos algunas de las representaciones más populares en el contexto de los hipergrafos y los complejos simpliciales, centrándonos en las distintas extensiones de la matriz de adyacencia al caso de orden superior.

La matriz de incidencia es la manera clásica en la que las matemáticas han descrito relaciones entre dos clases de objetos. Fue introducido por Kirchhoff en 1847 para modelar matemáticamente circuitos eléctricos. Aquí definimos la matriz de incidencia de un grafo $G = (V, E)$ como una matriz $n \times m$ y entradas $I = \{I_{ia}\}$, donde n es el número de nodos y m es el número de aristas. La entrada I_{ia} en la fila i y la columna a es 1 si el nodo i y la arista a están incidentes, y 0 en caso contrario. La definición se puede extender fácilmente al caso de interacciones de orden superior, en el que a etiqueta el tipo más general de interacción, como vemos en los ejemplos de la figura 2a y la figura 2b.

Es interesante ver que la matriz de incidencia también se puede entender como la matriz de adyacencia de un grafo bipartito con dos conjuntos de nodos, uno de tamaño n y otro de tamaño m .

A partir de la matriz de incidencia, también podemos construir otra matriz que codifique completamente la conectividad del grafo: la matriz de adyacencia A . Dado que el producto de matrices $I * I^T$ es una matriz $n \times n$ cuyo elemento (i,j) es el número de columnas de la matriz de incidencia I que contiene a los

vértices i y j , mientras que el elemento (i,i) da el grado del nodo i , la matriz de adyacencia de un grafo simple se puede definir como:

$$A = II^T - D \quad (1)$$

donde D es la matriz diagonal cuyas entradas en la diagonal son los grados de los nodos. La matriz de adyacencia A es 0 a lo largo de la diagonal, mientras que para $i \neq j$ la entrada $A_{ij} = 1$ si los nodos i y j son adyacentes, es decir, existe una arista que los conecta.

Podemos generalizar la noción de matriz de adyacencia al caso de hipergrafos y complejos simpliciales mediante al menos dos formas básicas:

- Usando la misma expresión que la ecuación 1 y considerando como D la matriz diagonal cuyas entradas en la diagonal son el número de hiperaristas a las que pertenece un vértice.

Es importante ver que para los grafos simples puede haber como máximo una arista que conecte a un par de nodos i y j , pero en hipergrafos puede existir más de una hiperarista α que contenga los dos nodos. La matriz de adyacencia de un hipergrafo es entonces una matriz $n \times n$ cuyo elemento A_{ij} captura el número de hiperaristas que contienen a los nodos i y j . Es interesante mencionar que, en el caso de que las hiperaristas sean ponderadas (pesadas), se puede extender el resultado anterior usando $A = IW^T - D$, donde W es una matriz diagonal con los pesos de las hiperaristas y D contiene el grado de los nodos.

- Se puede definir directamente, sin depender de la definición de una matriz de incidencia. Este enfoque se utiliza a menudo cuando las estructuras de orden superior no pueden ser identificadas de forma única solo por el conjunto de nodos implicados, o cuando hay una necesidad teórica de tener una noción más restrictiva de adyacencia entre dos *hyperedges*, más allá de que se cruzan en al menos dos vértices.

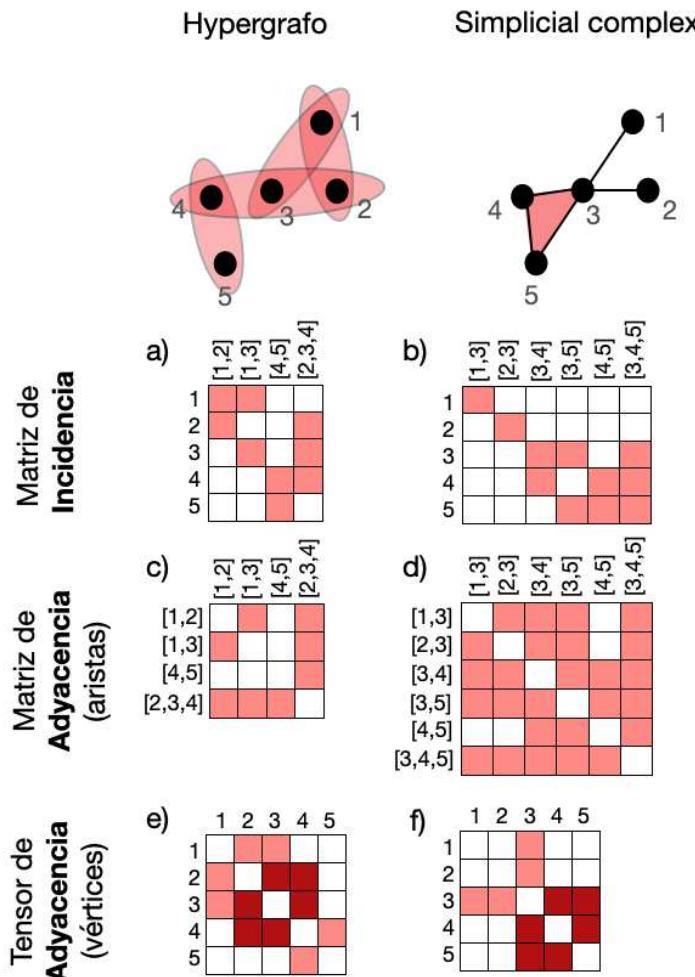
Por ejemplo, cuando se estudian motivos en una red (como pueden ser aristas o triángulos), para cada motivo M se puede construir una matriz de adyacencia $n \times n$, donde n es de nuevo el número de nodos, y cuyas entradas A_{ij} son el número de veces que i y j pertenecen a una instancia del motivo M .

En el caso con más interés, donde los motivos M son d -cliques (es decir, una arista, un triángulo, un tetraedro, etc.), entonces la matriz A_M se puede representar como un tensor de dimensión d , indexado por las etiquetas de los nodos.

Por ejemplo, consideramos hipergrafos con interacciones a pares y a triángulos, $A_M = \{A, B\}$, donde A es una matriz de adyacencia 1-simplex (la matriz del grafo simple) y B es un tensor de adyacencia 2-simplex (tensor de rango 3). Si asumimos que la red es no ponderada y no dirigida, tenemos que $A_{ij} = A_{ji} = 1$ si y solo si existe una arista entre i y j y $B_{ijl} = B_{ilj} = B_{jil} = B_{jli} = B_{lji} = B_{lij} = 1$ si y solo si existe una interacción a tres bandas entre los nodos i , j y l . La estructura de A y B puede estar en general sin correlacionar en un hipergrafo, pero en un complejo simplicial tenemos la restricción $B_{ijl} = A_{ij}A_{jl}A_{li}$, es decir, que solo existe un triángulo si existen las tres aristas que lo forman.

En la figura 2 observamos distintos ejemplos de representación. Los valores del tensor de adyacencia son agregados entre las distintas dimensiones del tensor. El color intenso indica la presencia de una interacción a tres cuerpos. Vemos que, mientras que en caso del hipergrafo, donde esa interacción $[2,3,4]$ con $B_{234} = 1$ no se sobrepone con ninguna arista en ninguna combinación de pares del triángulo, en el caso del complejo simplicial las entradas $(3,4)$ y $(4,5)$ tienen contribuciones tanto de la matriz de grafo A como del tensor de triángulos B . Para simplificar la visualización, hemos considerado $A'_{ij} = A_{ij} + \sum_l B_{ijl}$.

Figura 2. Distintas representaciones matriz/tensor de hipergrafos y complejos simpliciales



Ejemplo de representaciones

La figura 2 muestra distintas representaciones de matriz y/o tensor para el caso de un hipergrafo general (columna de la izquierda) o un complejo simplicial (columna de la derecha).

En particular, usamos la matriz de incidencia (a,b), la matriz de adyacencia en el espacio de las hiperaristas (c,d) y el tensor de adyacencia agregado en todas sus dimensiones (e,f).

2.3. Métricas de centralidad

Es interesante preguntarse cómo las métricas de centralidad de grafos vistas anteriormente pueden extenderse a los hipergrafos y a los complejos simpliciales.

Primero consideramos el grado de los nodos, que puede ser definido de manera directa como la suma de los elementos de la fila i de la matriz de incidencia:

$$\deg(v_i) = \sum_i^n I_{ia}, \quad (2)$$

y nos informa del número de hiperaristas en los que cada nodo contribuye (asumiendo que no están ponderadas o pesadas). Es importante ver que, en un grafo simple, la columna de una matriz de incidencia siempre suma 2, ya que las relaciones siempre son entre grupos de dos nodos. En cambio, en un hipergrafo o complejo simplicial, la suma de los elementos de las columnas de I define la secuencia de tamaños de los hipergrafos en el sistema.

De forma alternativa, podemos definir el grado de los nodos usando la representación tensorial de la matriz de adyacencia. En este caso, tenemos un grado distinto para cada tipo de hiperarista. Considerando solo 1 y 2-simplex (aristas y triángulos), podemos calcular directamente:

$$\deg_1(v_i) = \sum_j^n A_{ij} \quad (3)$$

$$\deg_2(v_i) = \frac{1}{2} \sum_j^n \sum_k^n B_{ijk}, \quad (4)$$

expresiones que nos informan del número de aristas y triángulos a los que cada nodo es incidente, respectivamente.

Ahora nos movemos del ámbito local en el que opera el grado hasta el ámbito global del hipergrafo. Para obtener una centralidad relativa a toda la red, necesitamos definir una manera clara en la cual las hiperaristas pueden ser atravesadas por caminos aleatorios o *random walks*.

Anteriormente, vimos que un camino de m pasos en un grafo simple es una secuencia de vértices, $w = \{v_1, v_2, \dots, v_m\}$, de tal manera que los vértices consecutivos son adyacentes en la red. En el caso de los hipergrafos, en general tenemos que considerar caminos que conectan parejas de hiperaristas en lugar de vértices, siendo el caso más sencillo el de un camino aleatorio definido como una secuencia de hiperaristas conectadas al menos a través de un vértice. En este caso, podemos definir la centralidad de caminos aleatorios del vértice v_i como el número de caminos cerrados de distinta longitud que empiezan y terminan en ese vértice. Usando las propiedades de la matriz de adyacencia y sus potencias, podemos contar el número de caminos cerrados de longitud k mediante:

$$RW_k(v_i) = (A^k)_{ii} = \sum_{s=1}^n (u_{is})^2 \lambda_s^k, \quad (5)$$

con λ_s el valor propio s de la matriz de adyacencia y U la matriz con los autovectores en columnas. Una posible métrica de centralidad de caminos aleatorios en hipergrafos (CRW por sus siglas en inglés) se obtiene sumando todos los caminos de distintas longitudes k , normalizando cada término por el factorial de $k!$ para pesar con mayor contribución a los caminos más cortos.

En particular, podemos calcular la métrica de centralidad usando la expresión:

$$CRW(v_i) = \sum_{k=1}^{\infty} \frac{RW_k(i)}{k!} \leq e^{\lambda_{max}}, \quad (6)$$

y acotarla superiormente por la exponencial del autovalor λ_{max} usando la definición de la función exponencial.

Es importante mencionar que existen otras medidas de centralidad global. Por ejemplo, la métrica conocida como *eigenvector centrality*, que cuenta caminos de longitud infinita, se puede extender a hipergrafos calculando el vector propio asociado al mayor valor propio de la matriz de incidencia del hipergrafo, obteniendo una medida de centralidad para todos los vértices y también para todas las hiperaristas, dadas por las componentes del primer vector propio. Por otro lado, se puede usar la representación tensorial o basada en motivos M para calcular la *eigenvector centrality* para cada motivo.¹

Para profundizar en los conceptos anteriores y descubrir otras medidas y métricas de utilidad en hipergrafos, recomendamos la lectura de la bibliografía complementaria.

Lectura complementaria

E. Estrada; J. A. Rodríguez-Velázquez (2005). «Complex networks as hypergraphs». arXiv:physics/0505137.

⁽¹⁾En el caso de dimensión 1, recuperaríamos la centralidad del grafo simple de aristas, y en el caso de dimensión 2, obtendríamos una centralidad basada en triángulos.

3. Aplicaciones

En el apartado anterior hemos visto una introducción a algunos de los aspectos teóricos más relevantes de las redes con interacciones de orden superior, centrándonos en definir y comprender los aspectos estructurales clave de este tipo de grafos más generales. A nivel formal, hemos visto como algunas de las definiciones y métricas conocidas en grafos simples pueden ser extendidas a estos casos, aunque no siempre existe una extensión natural y única y hay que tomar distintas decisiones en el proceso.

El desarrollo y estudio teórico de estos objetos no está principalmente motivado por el mero interés o curiosidad matemática en describir sus propiedades, sino que existe un gran número de aplicaciones, con creciente interés, donde este tipo de estructuras juegan un papel importante.

En este apartado, introduciremos los conceptos básicos y algunas referencias importantes sobre las aplicaciones de estos grafos. Primero, en el contexto del análisis topológico de datos, donde veremos como un método conocido como **homología persistente** permite extraer la forma y la geometría de un conjunto de datos mediante un proceso de filtraje y conversión de los datos a redes geométricas con interacciones de orden superior, y segundo, en el contexto del modelaje de **procesos dinámicos** que ocurren en sistemas reales interconectados a través de grafos, donde veremos como la presencia de estas interacciones puede inducir transiciones abruptas en la red.

3.1. Estructura: modelos geométricos y análisis topológico de datos

Primero nos centraremos en discutir algunas de las aplicaciones que explotan la estructura de los hipergrafos para entender la geometría de un conjunto de datos.

La idea de aproximar la estructura geométrica de un conjunto de datos multidimensional usando grafos simples ha tenido un éxito considerable en aplicaciones diversas, en ámbitos como la sociología computacional o la neurociencia. Por ejemplo, el algoritmo *Isomap*, referenciado como lectura complementaria, asocia un punto en el espacio a un vértice de un grafo y asume que dos vértices están conectados por una arista si los nodos están a menos de cierta distancia entre ellos. Analizando el grafo resultante, se pueden definir un nuevo conjunto de distancias entre los puntos y asignar un nuevo conjunto

Lectura complementaria

J. B. Tenenbaum; V. de Silva; J. C. Langford (2000). «A Global Geometric Framework for Nonlinear Dimensionality Reduction». *Science* (vol. 290, págs. 2319–2323).

de coordenadas de menor dimensión a cada punto, lo que permite revelar la geometría subyacente de los datos, así como encontrar comunidades o medir efectos de correlaciones locales en el grafo.

Es razonable pensar que estas ideas se pueden extender y mejorar usando hipergrafos. Por ejemplo, el método conocido como **homología persistente** usa hipergrafos y no se limita a fijar un umbral y trabajar con un solo grafo, sino que construye una secuencia de complejos simpliciales representados en un cierto **espacio geométrico**, dependiendo de un parámetro de filtración, siendo el más usual el radio de interacción entre nodos en el espacio geométrico. A medida que la filtración avanza, distintas propiedades topológicas de los datos son reveladas a partir de medidas estructurales del complejo simplicial, como por ejemplo el número de componentes conexos de cada dimensión (arestas, triángulos, etc.), y los bucles o agujeros del conjunto de simplex. La práctica usual consiste en monitorizar cómo las propiedades topológicas del complejo simplicial persisten a lo largo del proceso de filtración (desde un complejo simplicial constituido por vértices aislados hasta un simplex completo de dimensión $n - 1$) para descubrir patrones geométricos en los datos a distintas escalas de resolución.

Aquí no entraremos a explicar los detalles de este método, que requiere unos conocimientos avanzados de topología algebraica y sus definiciones. No obstante, es importante conocer la existencia y el potencial del método para analizar conjuntos de datos en un espacio multidimensional. Sugerimos la lectura complementaria que profundiza en el método, que presenta un tutorial de uso sencillo e ilustrativo.

Ejemplo de modelos geométricos de complejos simpliciales

Estos modelos son usados en el proceso de filtración de complejos simpliciales en el método de análisis topológico de datos conocido como **homología persistente**. Normalmente, la filtración se controla tuneando el parámetro r y se mide la persistencia de propiedades topológicas como el número de componentes conexos a distintos órdenes de interacción en función de r .

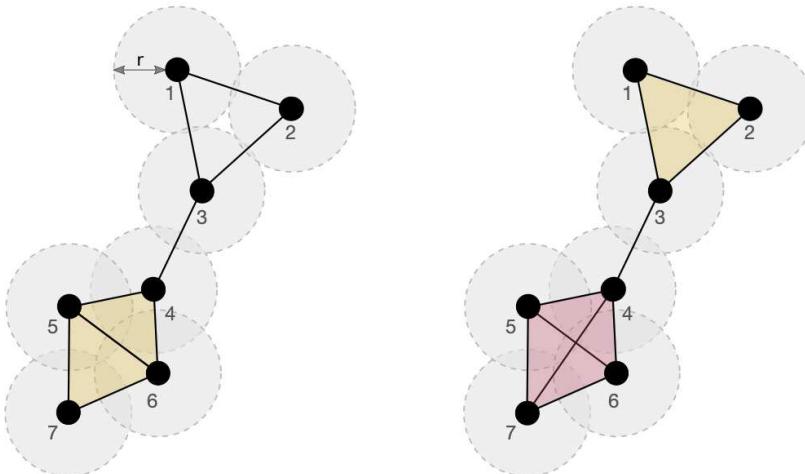
Un paso crucial en el análisis topológico de los datos y en homología persistente consiste en ubicar el conjunto de datos en un cierto espacio geométrico donde se pueden definir distancias entre puntos y construir el complejo simplicial a partir de estas distancias. En este contexto, son particularmente interesantes los modelos geométricos generativos. En la figura 3 vemos dos ejemplos relevantes. En el *Cech complex*, se construye un $(m - 1)$ -simplex entre m vértices si hay intersección de las m bolas de radio r . En el *Vietoris-Rips complex*, la restricción es menor y se construye el simplex si existen los simplex de orden inferior.

Lectura complementaria

N. Otter; M. A. Porter; U. Tillmann; P. Grindrod; H. A. Harrington (2017). «A roadmap for the computation of persistent homology». *EPJ Data Science* (págs. 6-17). DOI 10.1140/epjds/s13688-017-0109-5.

Figura 3. Dos ejemplos clave de modelos geométricos de complejos simpliciales

a) Cech complex b) Vietoris-Rips complex



Por otro lado, cuando los datos o vértices son distribuidos de forma aleatoria en el espacio, estos **modelos geométricos** sencillos permiten descubrir clasificaciones muy interesantes de los comportamientos de los distintos tipos de complejos simpliciales en función de los parámetros de filtración. Estos resultados son de naturaleza muy matemática y van más allá de los contenidos propuestos para este material, pero sugerimos la lectura complementaria al lector interesado.

**Lectura
complementaria**

O. Bobrowski; M. Kahle (2018). «Topology of random geometric complexes: a survey». *J. Appl. Comput. Topol.* (vol. 1, n.º 3-4, págs. 331-364).

Ejemplo de modelos geométricos

La figura 3 muestra dos ejemplos clave de modelos geométricos de complejos simpliciales.

En concreto, en la figura 3b, vemos que se ha formado un triángulo entre los vértices (1,2,3) porque existe intersección entre los tres vértices por pares, y que se ha formado un tetraedro entre los vértices (4,5,6,7) porque existe intersección de triángulos.

En cambio, en la figura 3a, vemos que, con el mismo radio de interacción, solo se han formado el par de triángulos (4,5,6) y (5,6,7). Dependiendo de la aplicación, puede ser más interesante usar un modelo u otro, aunque es importante tener en cuenta que las propiedades del análisis topológico en distintas filtraciones (variando r) puede depender de forma significativa de la elección del modelo.

Para concluir este subapartado, es interesante mencionar otro enfoque basado en **modelos probabilísticos** de hipergrafos. Si los modelos geométricos son útiles para el análisis topológico de los datos, los modelos probabilísticos son útiles para construir modelos nulos y poder medir la significancia estadística de ciertos patrones y motivos en hipergrafos reales. Por ejemplo, en un hipergrafo de coautores de artículos científicos o de colaboraciones de actores en películas, puede ser interesante entender si existen efectos de modularidad, *clustering* o propiedades particulares de los complejos simpliciales que no se pueden explicar como consecuencia del azar, lo que indicaría que existen mecanismos generativos detrás de los patrones observados. Podemos exten-

der los modelos clásicos de grafos como el modelo de Erdős-Rényi, el modelo de **configuración** o de **grafos exponenciales aleatorios** o el modelo de **bloques estocásticos** al caso con interacciones de orden superior. Estos modelos se consideran nulos (es decir, aleatorios y sujetos a ciertas restricciones, por ejemplo, el número de nodos, el número de hiperaristas o la secuencia de grados en los nodos), porque permiten comprender mejor las métricas y propiedades peculiares de hipergrafos reales con respecto a sus hipergrafos aleatorios asociados.

La descripción y el análisis de estos modelos probabilísticos en el contexto de los hipergrafos adquiere rápidamente una alta complejidad técnica debido al crecimiento exponencial de los patrones combinatorios en hipergrafos de creciente dimensión. Los detalles se pueden encontrar en las referencias de la bibliografía, pero su estudio queda más allá de los contenidos de este material didáctico. No obstante, es importante entender que los modelos expuestos anteriormente en el caso de grafos simples se pueden extender de forma natural a hipergrafos para capturar e interpretar mejor datos relacionales complejos.

3.2. Dinámica: efectos abruptos en transiciones de fase

Una aplicación muy relevante de las redes con interacciones de orden superior es la incorporación y análisis de estas estructuras en los modelos matemáticos que describen procesos reales. Es bien conocido que las interacciones más allá de por pares están presentes en muchos sistemas reales, tanto sociales como biológicos y tecnológicos. Por ejemplo, en procesos de contagio “complejo” de opiniones, la presión de todo el grupo (y no solo las interacciones entre parejas) puede influenciar de manera significativa en la adopción de una opinión o la propagación de una idea en la población. También en el contexto de redes neuronales o ecológicas, las interacciones a triángulos (a tres cuerpos) son necesarias para describir con precisión los procesos que observamos en estas redes.

La creciente evidencia acerca de la ubicuidad y relevancia de estas interacciones complejas, más allá de por pares, en sistemas reales está discutida en profundidad en la referencia complementaria y en la bibliografía del material didáctico. En cualquier caso, es importante mencionar que aún están poco entendidos los efectos de estas interacciones en la dinámica macroscópica o en las variables observables de los procesos en grafos, y hay muchos trabajos recientes que siguen esta interesante línea de investigación.

Lectura complementaria

F. Battiston y otros (2021). «The physics of higher-order interactions in complex systems». *Nature Physics* (vol. 17, págs. 1093–1098) [1].

A continuación, explicaremos de forma breve cómo introducir interacciones de orden superior en los modelos dinámicos, y veremos alguno de los efectos más importantes inducidos por estas interacciones en la dinámica macroscópica.

De forma bastante general, podemos definir un cierto proceso dinámico que ocurre en una red, considerado unidimensional por simplicidad, como una ecuación diferencial para el cambio temporal de la variable de interés:

$$\frac{dx_i}{dt} = F_i(x_i) + K_1 \sum_{j=1}^n A_{ij} G_i^{(2)}(x_i, x_j) + K_2 \sum_{j,k=1}^n B_{ijk} G_i^{(3)}(x_i, x_j, x_k) + \dots, \quad (7)$$

para cada vértice v_i de un grafo. La variación de la cantidad x_i en el tiempo, en cada vértice, viene determinada por un primer término $F_i(x_i)$ que captura la dinámica intrínseca del vértice (por ejemplo, la frecuencia de un oscilador o la opinión de un agente). El primer término de interacción, controlado por la constante de interacción K_1 , captura la interacción usual por pares mediante una función de acoplamiento $G_i^{(2)}(x_i, x_j)$, que en general depende de los estados de pares de nodos, en los cuales hay una conexión, es decir, donde la entrada de la matriz de adyacencia $A_{ij} = 1$. El segundo término de interacción está controlado por la constante de interacción K_2 , que captura interacciones a tres cuerpos, es decir, donde la función de interacción $G_i^{(3)}(x_i, x_j, x_k)$ depende de las variables de tres nodos conectados en un «triángulo», para el cual la entrada del tensor de adyacencia (en este caso un tensor de rango 3, con 3 dimensiones) es $B_{ijk} = 1$. El formalismo expuesto anteriormente permite introducir interacciones más allá de tres cuerpos, considerando interacciones a cuatro, cinco cuerpos, etc. Aquí nos centraremos en el caso más simple, donde estudiamos el efecto de interacciones a pares y a triángulos (es decir, solo considerando interacciones a dos y a tres cuerpos).

En el contexto de la propagación de una idea o un virus biológico, x_i se puede entender como la probabilidad de que el agente en el vértice v_i esté en un comportamiento dado, por ejemplo infectado o susceptible, y más allá de modelar la interacción entre pares de agentes mediante un parámetro de infectividad (representado por K_1), se asume que hay también un efecto de grupo (o de triángulos en este caso concreto), que afecta la variación de la probabilidad de estado de un nodo. En particular, K_2 representa la probabilidad de infección del vértice v_i en el caso de que los nodos v_j como v_k estén infectados. En cambio, en el contexto de la dinámica de osciladores acoplados, x_i puede ser entendido como la fase del oscilador, θ_i , la función a pares $G_i^{(2)} = \sin(\theta_j - \theta_i)$ y la función a tres cuerpos puede ser definida como $G_i^{(3)} = \sin(\theta_j + \theta_k - 2\theta_i)$, aunque existen otras funciones que producen un efecto similar. De hecho, es importante mencionar que, al añadir interacciones a tres cuerpos, normalmente se abre un abanico más grande de posibilidades para construir las funciones de interacción.

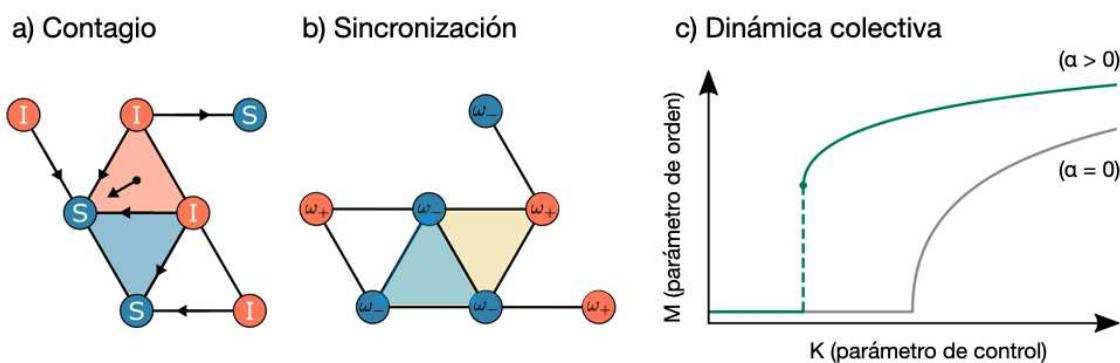
La simulación de la dinámica propuesta en la ecuación 7 se puede implementar de manera sencilla usando cualquier método de integración de ecuaciones diferenciales (Euler, Heun, Runge-Kutta, etc.), dadas unas condiciones iniciales $x(t=0)$ y una estructura de grafo conocida, tanto en su matriz de adyacencia a pares, $A \in R^{n \times n}$ como su tensor de adyacencia que captura interacciones a

Lectura complementaria

I. Iacopini; G. Petri; A. Barrat; V. Latora (2019). «Simplicial models of social contagion». *Nature Communications* (vol. 10, n.º 2485). DOI: 10.1038/s41467-019-10431-6 [2].

triángulos, $B \in \mathbb{R}^{n \times n \times n}$. Los resultados que discutiremos a continuación, que se recogen de manera detallada en las referencias complementarias de este subapartado, consideran la simulación numérica del sistema descrito en la ecuación 7 para el caso de la propagación en el modelo SIS y en el modelo de Kuramoto de osciladores acoplados, con las interacciones a tres cuerpos descritas con las funciones presentadas en el párrafo anterior. Para la estructura, consideramos el caso de un complejo simplicial, donde los triángulos $B_{ijk} = 1$ vienen delimitados por la presencia de las tres aristas en la matriz A , es decir, $B_{ijk} = A_{ij}A_{jk}A_{ki}$.

Figura 4. Efecto de las interacciones a tres cuerpos en las transiciones epidémicas o de sincronización



Fuente: Bick y otros (2021)

Efecto de las interacciones a tres cuerpos en las transiciones epidémicas o de sincronización

La figura 4a muestra un ejemplo de cómo la propagación de un virus puede ser mediada por interacciones a pares o a triángulos; en la figura 4b vemos la misma idea en el contexto de osciladores heterogéneos (con frecuencias ω_+ y ω_-), y en c) vemos cómo la activación de las interacciones a tres cuerpos $\alpha > 0$ puede hacer cambiar la naturaleza de la transición dinámica, volviéndola abrupta en el parámetro de orden M (asumido en promedio para tiempos largos), que representa o bien la prevalencia de infectividad o el grado de sincronización global. Este fenómeno puede explicar el origen de las transiciones abruptas en fenómenos reales de propagación de ideas o las transiciones explosivas de sincronización que ocurren en el cerebro, por ejemplo, en la transición de conciencia-inconciencia al dormirse o en la sincronización abrupta de regiones cerebrales que ocurre en un episodio de ataque epiléptico. Figura adaptada de la referencia (Bick y otros, 2021).

Para entender el resultado principal de estos trabajos, resumido en la figura 4, definimos la intensidad o probabilidad de interacción a pares como $K_1 = K$ y la interacción a triángulos como $K_2 = \alpha K$, de manera que $\alpha > 0$ nos regula la contribución relativa de las interacciones a triángulos. El parámetro de orden en el caso de una epidemia es simplemente $M(t) = \langle x(t) \rangle$, donde $\langle x(t) \rangle$ es la media de la probabilidad de infección en todos los vértices. En el caso de considerar un sistema de osciladores acoplados $M(t) = N^{-1} |\sum_j e^{i\theta_j}|$, es decir, el módulo del parámetro de orden que mide el grado de sincronización global en el sistema entre cero y uno.

En la figura 4c, vemos que tanto en el modelo de propagación de epidemias como en el modelo de osciladores acoplados, para un cierto $\alpha > 0$, la tran-

Lectura complementaria

P. S. Skardal; A. Arenas (2019). «*Abrupt desynchronization and extensive multistability in globally coupled oscillator simplexes*». *Physical Review Letters* (vol. 122, pág. 248301). [3]

sición entre el estado no endémico (o incoherente) y el estado endémico (o sincronizado) pasa de ser una transición de segundo orden en K (es decir, que la transición es suave –no abrupta– en el parámetro de orden M a medida que aumenta el parámetro de control $K > K_c$) a ser una transición de primer orden –abrupta– (donde el parámetro de orden cambia de forma abrupta para un cierto valor crítico K_c).

Además, en los dos trabajos discutidos anteriormente, se descubrió que esta fenomenología observada en los dos modelos es robusta, es decir, que se mantiene para distintas elecciones de las funciones de interacción y del tipo de hipergrafo, señalando que la emergencia de transiciones abruptas se puede explicar mediante la inclusión de interacciones a tres cuerpos. Este hecho apunta a la existencia de un formalismo general para modelar y entender el origen de este tipo de transiciones abruptas que se observan en la naturaleza y en la sociedad. Es importante mencionar que, a pesar de esta importante evidencia preliminar, no existe una prueba rigurosa de este fenómeno y de su robustez en distintas dinámicas y tipos de grafos, y esta línea de investigación sigue siendo explorada en la actualidad. En este contexto, es interesante mencionar un trabajo reciente de Kuehn y Bick en 2021, referenciado como lectura complementaria, que propone un argumento general basado en la teoría matemática de bifurcaciones para explicar como variaciones en modelos de interacción a pares (por ejemplo, activando las interacciones a tres cuerpos con $\alpha > 0$), pueden llevar a un cambio del comportamiento crítico (de continuo a discontinuo), en una clase muy amplia de modelos, incluyendo la sincronización y la propagación de epidemias.

Lectura complementaria

Kuehn, C.; Bick, C. (2021). «A universal route to explosive phenomena». *Science Advances* (vol. 7, n.º 16). DOI: 10.1126/sciadv.abe3824.

Resumen

En este material didáctico hemos presentado los conceptos necesarios para entender las redes con interacciones de orden superior, los llamados hipergrafos y una clase particular de ellos, los complejos simpliciales.

En el apartado 2 hemos visto algunas de las limitaciones de los grafos simples para explicar interacciones en sistemas sociales y hemos motivado el uso de los hipergrafos en el análisis de datos sociales.

El apartado 3 de este material ha introducido el marco teórico básico para comprender las matemáticas que describen estos objetos. Hemos visto las definiciones formales de hipergrafos y complejos simpliciales, sus distintas representaciones en forma matricial y algunas de las extensiones de las métricas de centralidad vistas anteriormente.

Una vez comprendidos los conceptos teóricos principales, hemos podido introducir algunas aplicaciones de los hipergrafos en el análisis de datos y la modelización de procesos físicos. En particular, en el apartado 4, hemos visto aplicaciones basadas en explotar la estructura de los datos mediante modelos geométricos de hipergrafos y su análisis topológico y hemos discutido el efecto de las interacciones entre tres nodos en los procesos dinámicos de propagación de información en redes, descubriendo que estas interacciones superiores pueden explicar las transiciones explosivas o abruptas en las variables macroscópicas (es decir, en la dinámica colectiva) observadas en distintos sistemas sociales y naturales.

En conclusión, los contenidos presentados en las páginas anteriores ofrecen el conocimiento y las herramientas necesarias para modelar y analizar las redes con interacciones de orden superior y para entender el papel que juegan los hipergrafos y los complejos simpliciales en la descripción de los datos relacionales en sistemas sociales y en los procesos físicos que ocurren en redes reales, como la propagación de información en la sociedad o la sincronización de osciladores acoplados en el cerebro.

Glosario

análisis topológico *m* Rama de la matemática dedicada al estudio de aquellas propiedades de los cuerpos geométricos que permanecen inalteradas por transformaciones continuas.

matriz laplaciana *f* Representación matricial de un grafo diseñada para capturar las propiedades espectrales de los grafos. También se conoce como matriz de admitancia o matriz de Kirchhoff.

Bibliografía

Barabási, A.-L.; Pósfai, M. (2016). *Network science*. Cambridge: Cambridge University Press. ISBN: 9781107076266 1107076269.

Battiston, F.; Cencetti, G.; Iacopini, I.; Latora, V.; Lucas, M.; Patania, A.; Young, J. G.; Petri, G. (2020). «Networks beyond pairwise interactions: Structure and dynamics». *Physics Reports* (vol. 874, págs. 1-92).

Bianconi, G. (2021). *Higher-Order Networks. An Introduction to Simplicial Complexes* [publicado en línea]. Cambridge University Press. ISBN: 9781108770996.

Bick, C.; Gross, E.; Harrington, H. A.; Schaub, M. T. (2021). *What are higher-order networks?* ArXiv repository: arXiv:2104.11329.

Bretto, A. (2013). «Hypergraph Theory: An Introduction». *Mathematical Engineering, Springer Science and Business Media*. ISBN: 3319000802, 9783319000800.

Newman, M. (2018). *Networks (Second Edition)*. Oxford University Press. ISBN: 9780198805090.