UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Clasificaciones deportivas y búsquedas en internet

Asomándote a las redes complejas

Apellidos, nombre	Carpitella, Silvia[1] (carpitella@utia.cas.cz) Izquierdo Sebastián, Joaquín[2] (jizquier@upv.es)
Departamento	[1] Department of Decision-Making Theory [2] Departamento de Matemática Aplicada
Centro	[1] Institute of Information Theory and Automation, Praga [2] Universitat Politècnica de València

1 Resumen

Recientemente se ha acuñado el término *Redes complejas* para referirse a una moderna rama de la Física que trata de modelar las interrelaciones entre los elementos de un conjunto finito de objetos con el fin de observar el comportamiento emergente resultante. A pesar del respeto que produce la palabra *complejidad*, en este artículo nos asomamos brevemente mediante algunos ejemplos que: i) interesarán al alumno y ii) utilizan técnicas algebraicas sencillas. Este artículo está pensado como una clase motivadora para el Álgebra Matricial, asignatura que forma parte del currículo en las Escuelas y Facultades universitarias. Aquí, se ha aplicado en la asignatura Álgebra del Grado en Ingeniería Física de la UPV. Los ejemplos considerados incluyen dos elementos motivadores: las clasificaciones deportivas y las búsquedas por internet. En el desarrollo, se muestra la necesidad de determinadas herramientas avanzadas (que forman parte esencial en la asignatura), lo que estimula el interés de aprender tales herramientas.

2 Introducción

Modelar las interrelaciones entre un conjunto finito de objetos es de gran interés en muchos problemas de diversos tipos. Por ejemplo, para describir varios tipos de redes (carreteras que conectan ciudades, rutas de aerolíneas que conectan aeropuertos, redes de comunicación que conectan satélites, etc.); también, relaciones entre grupos o individuos (relaciones de amistad en una sociedad, relaciones depredador-presa en un ecosistema, relaciones políticas y/o comerciales entre naciones o entre alianzas de ellas, clasificación en un deporte de competición, etc.).

Los grafos son herramientas adecuados para modelar tales redes y relaciones, y **las matrices** son una herramienta útil en su estudio.

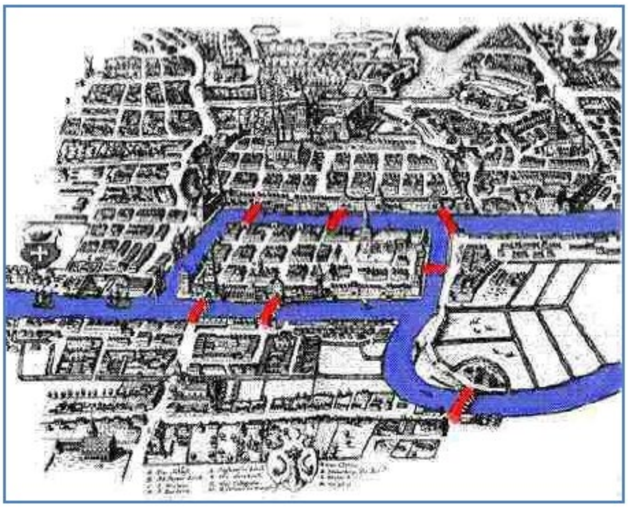

Figura 1. El problema de los puentes de Könisberg [Fuente: Johnson (2017)]

El término grafo fue empleado por primera vez a mediados del siglo XVIII y se originó a partir de la "*notación gráfica*" que entonces, y aún en la actualidad, permite describir los enlaces moleculares en el ámbito de la química orgánica/bioquímica. El origen de la teoría de grafos como ciencia se da en el mismo siglo, ubicándose en la antigua Könisberg (actualmente Kaliningrado) en Rusia (ver Figura 1). Leonard Euler, publicó un documento en el que se le daba solución al llamado *Problema de Könisberg*, que perseguía encontrar una manera de recorrer los siete puentes de la ciudad pasando por cada uno de ellos una sola vez. Posteriormente, en el año 1847, Gustav Kirchoff, produce una serie de publicaciones, en la que hace uso de la teoría de grafos para establecer las ampliamente conocidas leyes para cálculo de corriente, voltaje y resistencia en los circuitos eléctricos.

Actualmente la teoría de grafos es esencial en la moderna rama de la Física que se ha venido en denominar *Redes complejas*.

En este documento encontrarás **en rojo** conceptos que quizá no conozcas y que vamos a ver en la asignatura de Álgebra; y **en azul** encontrarás conceptos que seguramente ya conoces, pero que deberías refrescar (algunos los vemos también en la asignatura).

3 Objetivos

Tras concluir con la lectura de este documento, serás capaz de:

- Describir algunas redes complejas sabiendo indicar razones de su complejidad.
- Explicar con detalle los elementos clave de la complejidad.
- Obtener soluciones de los modelos planteados.

4 Desarrollo

Para leer este artículo deberás tener nociones de:

Requisitos
1. Álgebra matricial elemental.
2. Solución de sistemas de ecuaciones lineales.
3. Operaciones simples en algún entorno computacional.

Tabla 1. Requisitos básicos

5 Clasificaciones deportivas

Empecemos con un ejemplo muy fácil de entender. Observa el grafo siguiente.

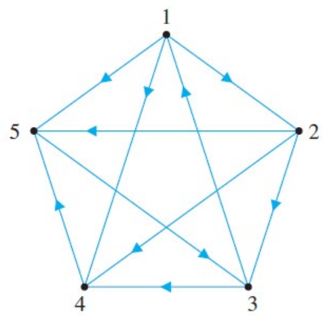

Figura 2. Grafo dirigido (resultados de un torneo entre 5 jugadores)

Ejemplo 1. Un torneo es una competición en el que cada participante juega exactamente una vez contra los demás. La figura 2 es un **grafo dirigido** (a veces llamado *digrafo*) en el que los jugadores (o equipos) están representados por los puntos numerados (en teoría de grafos, se los denomina **vértices del grafo**); las líneas (se las llama **aristas**) que conectan los vértices representan los partidos jugados (observa que están conectados todos con todos, ya que se trata de un torneo); y una flecha sobre una arista que conecta un vértice con otro, indica una relación entre esos dos vértices; y en nuestro caso representa que el primer jugador ganó al segundo.

Para este grafo dirigido, su denominada **matriz de adyacencia** es

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix},$$

donde

- la numeración de los vértices marca el orden de las **filas** y de las **columnas** de la matriz, y
- un 1 en la **posición** (i, j) indica que el jugador i ganó al j.

Por ejemplo, el 1 último de la primera fila, $a_{15} = 1$, indica que el jugador 1 (primera fila) ganó al 5 (quinta columna) en el partido entre ellos; esto está obviamente relacionado con el hecho de que hay una flecha que va del 1 al 5.

Ejercicio 5.1. Comprueba con detalle la correspondencia entre las aristas del grafo y las **entradas o elementos de la matriz** de adyacencia.

5.1 Clasificación

Supongamos que queremos clasificar a estos jugadores de acuerdo con los resultados del torneo. Una manera inmediata consistirá en sumar las victorias de cada uno de ellos y comparar los resultados. Esto es sencillo: fíjate que el número de victorias de cada jugador es el número de flechas que salen de su vértice y esto es, justamente, la suma de los elementos de su fila en la matriz de adyacencia. Sencillo, ¿verdad?

Ejercicio 5.2. Calcula el número de victorias de cada jugador según se dice en el párrafo anterior a partir de a) las aristas del grafo; b) los 1s que tiene cada fila.

No obstante, interesa ver esto desde otro punto de vista (que, quizá, te parezca extravagante): el número de victorias de cada jugador se obtiene **multiplicando la matriz A por el vector**

$$\mathbf{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Pues, en efecto, se tiene que

$$A \cdot \mathbf{u} = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 1 \end{pmatrix},$$

lo que clasifica en primer lugar, empatados, a los jugadores 1 y 2; luego tercero queda el jugador número 3; y cierran la clasificación (también empatados) el 4 y el 5.

Ejercicio 5.3. Realiza el producto $A \cdot \mathbf{u}$. ¡Tienes que recordar cómo se hace! Y observa cómo el resultado contiene, en orden, la suma de los elementos de cada fila. ¿Coincide el resultado con lo obtenido en el Ejercicio 5.2?

5.2 Mejorando la clasificación

Probablemente esta clasificación no es satisfactoria y habrá criterios múltiples para dirimir los desempates. Un criterio podría utilizar el número de victorias indirectas de cada jugador. Una **victoria indirecta** de A sobre C se produce cuando A ha ganado a algún jugador B, quien, a su vez, ha ganado a C. Esta idea es especialmente interesante porque, justamente, se puede ver que una victoria indirecta corresponde a una trayectoria en el grafo en dos etapas desde A hasta C a través del vértice B: una flecha va de A a B y otra va de B a C.

Por ejemplo, en la figura se puede ver que el 5 ganó al 3 (flecha de 5 a 3), quien, a su vez, ganó al 4 (flecha de 3 a 4). Así que el 5 tiene (al menos) esta victoria indirecta en su haber. Y, de manera especialmente interesante, también se puede demostrar que el número de victorias indirectas se puede obtener fácilmente a través de la matriz A^2, el **cuadrado de la matriz** A (que no es otra cosa que el **producto matricial** de A por sí misma). Esta matriz es

$$A^2 = \begin{pmatrix} 0 & 0 & 2 & 1 & 2 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Siguiendo con el ejemplo $5 \to 3 \to 4$, vemos que, en efecto, la posición $(5,4)$ de A^2 tiene un 1, lo que indica, que el jugador 5 no tiene más victorias indirectas sobre el 4, aunque tiene otra victoria indirecta sobre el 1 (a través del 3: mira bien el grafo).

Ejercicio 5.4. ¿Te atreves a calcular este cuadrado? Quizá lo estudiaste antes; si no, lo veremos en esta asignatura.

Para obtener la clasificación final de los jugadores vamos a tener en cuenta, pues, las victorias directas (mediante $A \cdot \mathbf{u}$) y las indirectas (mediante $A^2 \cdot \mathbf{u}$). Esto, utilizando la **propiedad distributiva del producto matricial**, se realiza mediante la operación

$$A \cdot \mathbf{u} + A^2 \cdot \mathbf{u} = (A + A^2) \cdot \mathbf{u} = \begin{pmatrix} 0 & 1 & 2 & 2 & 3 \\ 1 & 0 & 2 & 2 & 2 \\ 1 & 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 7 \\ 6 \\ 2 \\ 3 \end{pmatrix},$$

que clasifica (ahora sin empates, aunque esto no tiene por qué ser siempre así) a los jugadores en el orden: 1, 2, 3, 5, 4, con 8, 7, 6, 3 y 2 victorias (directas e indirectas), respectivamente.

Ejercicio 5.5. Realiza con detalle los cálculos de las victorias directas e indirectas.

5.3 Complejidad

Puedes ver que esto no resulta en exceso complicado. No obstante, puedes imaginar problemas más grandes, correspondientes a grafos con muchos vértices y aristas, y cómo se hace imprescindible disponer de una herramienta más eficiente. Vamos a ello.

Lo que buscamos es asociar a un jugador i una clasificación c_i de manera que $c_i > c_j$ indique que el jugador i está mejor clasificado que el j (obvio). Para poder manejar bien estas clasificaciones (observa que, si hay muchos elementos a clasificar, las sumas de 'victorias' de diversos tipos podrían hacerse muy grandes) vamos a hacer que las clasificaciones c_i sean números parecidos a las probabilidades, es decir, tales que $0 \leq c_i \leq 1$, y, además, $\sum_{i=1}^{n} c_i = 1$. Obviamente, i varía entre 1 y n, siendo n el número de competidores.

Para los $n = 5$ jugadores de nuestro ejemplo, un vector genérico de clasificaciones sería

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix},$$

siendo cada c_i el número asociado al jugador i, fácilmente comparable con los otros.

Además, utilicemos la siguiente lógica: la clasificación de un jugador i debe ser proporcional a la suma de las clasificaciones de los jugadores a los que ha vencido. Por ejemplo, el jugador 1 ha vencido a los jugadores 2, 4 y 5. Así que exigiremos

$$c_1 = \alpha(c_2 + c_4 + c_5),$$

siendo α un coeficiente de proporcionalidad (no nulo, claro) adecuado.

Exigiendo lo mismo para todos los jugadores, se tiene el siguiente **sistema de ecuaciones**

$$\begin{aligned} c_1 &= \alpha(c_2 + c_4 + c_5) \\ c_2 &= \alpha(c_3 + c_4 + c_5) \\ c_3 &= \alpha(c_1 + c_4) \\ c_4 &= \alpha c_5 \\ c_5 &= \alpha c_3 \end{aligned}.$$

Ejercicio 5.6. Comprueba estas ecuaciones con detalle.

Fíjate que este sistema de ecuaciones se puede escribir matricialmente como

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \alpha \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} \quad \text{es decir } \mathbf{c} = \alpha A \mathbf{c}.$$

Así, el vector de clasificaciones, $\mathbf{c}$, debe cumplir $\alpha^{-1}\mathbf{c} = A\mathbf{c}$ o, de forma matricial,

$$(A - \alpha^{-1} I)\mathbf{c} = 0,$$

siendo I la **matriz identidad** de tamaño 5×5.

Observa que es el **sistema homogéneo** de cinco ecuaciones con cinco incógnitas

$$\begin{pmatrix} -\alpha^{-1} & 1 & 0 & 1 & 1 \\ 0 & -\alpha^{-1} & 1 & 1 & 1 \\ 1 & 0 & -\alpha^{-1} & 1 & 0 \\ 0 & 0 & 0 & -\alpha^{-1} & 1 \\ 0 & 0 & 1 & 0 & -\alpha^{-1} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

En selectividad, sobre sistemas como este (quizá algo más pequeños), se preguntan cuestiones típicas como: hallar el valor de α para que este sistema tenga más soluciones que la trivial (recuerda que un sistema lineal homogéneo siempre tiene la **solución trivial**, pero justo esa no nos interesa en absoluto). Otra cuestión inmediata que se suele preguntar es: para los valores de α hallados, resolver el sistema.

En el caso de la clasificación deportiva que estamos considerando, de las infinitas soluciones que tiene este sistema para el o los valores α que lo hagan compatible (indeterminado), nos interesa alguna que cumpla $\sum_{i=1}^{5} c_i = 1$.

En la asignatura vamos a estudiar de manera general este tipo de problemas, dentro de la unidad dedicada a la **teoría espectral**.

Por cierto, para el problema de los cinco jugadores, se obtiene el vector

$$\mathbf{c} = \begin{pmatrix} 0.29 \\ 0.27 \\ 0.22 \\ 0.08 \\ 0.14 \end{pmatrix},$$

que clasifica a los jugadores en el siguiente orden: 1, 2, 3, 5, 4; ahora sin empates.

6 Búsquedas en internet

Ideas semejantes pueden servir para entender cómo funcionan las búsquedas por Internet, por ejemplo, mediante Google. Los buscadores antiguos devolvían los resultados de las búsquedas sin ordenar. Esto hacía que sitios realmente interesantes quedasen relegados a páginas posteriores de la búsqueda. Encontrar cosas de interés resultaba misión muy complicada y requería mucho esfuerzo. Actualmente, los buscadores, por ejemplo, Google, devuelven los resultados ordenados de acuerdo con alguna medida de su relevancia. Una ligera variación del método visto para clasificar jugadores puede servir para entender cómo Google clasifica las búsquedas. Salvo por ciertas modificaciones que optimizan el proceso, es lo que hace Google con su famoso algoritmo *PageRank*.

En lugar de jugadores, ahora tenemos páginas web (vértices del grafo) con diversos links (aristas del grafo) entre ellas. Veamos un ejemplo manejable a mano.

Ejemplo 2. Una pequeña intranet tiene cuatro webs. La estructura de enlaces entre ellas es la siguiente:

(1; 2,3); (2; 3,4); (3; 1,2); (4; 1,2).

Por ejemplo, (3; 1,2) significa que la web 3 tienen enlaces a la 1 y a la 2.

Obtenemos una matriz 4×4, M, asociada a dicha intranet mediante la siguiente regla:

$m_{ij} = 1$ si la web i es accesible desde la web j;

$m_{ij} = 0$, en caso contrario.

$$M = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Por ejemplo, el elemento $m_{14} = 1$ porque la web 1 es accesible desde la 4.

Ejercicio 6.1. Comprueba todas las entradas de la matriz M.

Se trata de una **matriz semipositiva** (todos sus elementos son no negativos, y hay alguno que es positivo, es decir, no todos son nulos).

El objetivo es expresar la navegación dentro de la intranet mediante la probabilidad de pasar de una página a otra (suponiendo que los enlaces en una página web tienen probabilidades iguales de ser clicados).

Al dividir los elementos de cada columna por la suma de la columna, se obtiene la matriz

$$A = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \end{pmatrix}.$$

Fíjate que ahora los elementos de la matriz son 'probabilidades', porque: 1) son valores comprendidos entre 0 y 1; y 2) tienen suma igual a 1 en cada columna.

Esta matriz, además de semipositiva, es, por tener las propiedades 1) y 2) del párrafo anterior, una **matriz estocástica**.

Como en el caso de la clasificación deportiva, queremos asignar valores de importancia a las distintas páginas web de tal manera que la importancia asignada a una página sea proporcional a la suma de las importancias de las páginas que la referencian.

Si denominamos por $\mathbf{v}$ el vector de tales valores de importancia, se puede ver que $\mathbf{v}$ verifica la ecuación

$$A\mathbf{v} = \lambda\mathbf{v}.$$

Esta ecuación se denomina **ecuación de valor y vector propios** y la estudiaremos en la unidad de **Teoría espectral**.

Como anteriormente, utilizando, la matriz identidad, I, el problema se puede escribir como

$$(A - \lambda I)\mathbf{v} = 0,$$

que, como en el caso de las competiciones, es un sistema lineal homogéneo que depende del parámetro λ y que tiene solución no trivial solo para ciertos valores de dicho parámetro, que hay que determinar (de nuevo el problema de selectividad antes mencionado).

Quizá recuerdes que, para que ese sistema sea **compatible indeterminado**, el **determinante** de la matriz de coeficientes, $A - \lambda I$, debe ser $|A - \lambda I| = 0$. Este determinante da un **polinomio**, $p(\lambda)$, de grado n (n es el **tamaño de la matriz** A) cuyas **raíces** hay que calcular para resolver la denominada **ecuación característica** de A, $p(\lambda) = |A - \lambda I| = 0$ y, así, encontrar los valores de λ para los que el sistema tiene soluciones interesantes (no triviales).

Dado que A es una **matriz estocástica regular**, cumple el **teorema de Perron**, y dicho polinomio tiene una raíz mayor (en valor absoluto) que las demás, y es $\lambda = 1$.

El siguiente paso es resolver el sistema para este valor de $\lambda = 1$, es decir $(A - I)\mathbf{v} = 0$, que, como hemos dicho, es compatible indeterminado.

Una vez más, de entre las infinitas soluciones que se obtengan, nos interesa aquella que cumpla que sus componentes sean probabilidades, es decir,

$$0 \leq v_i \leq 1 \quad \text{y} \quad \sum_{i=1}^{n} v_i = 1.$$

En el caso de la matriz A dada, de tamaño $n = 4$, la ecuación de valor y vector propios es

$$(A - \lambda I)\mathbf{v} = \begin{pmatrix} -\lambda & 0 & 1/2 & 1/2 \\ 1/2 & -\lambda & 1/2 & 1/2 \\ 1/2 & 1/2 & -\lambda & 0 \\ 0 & 1/2 & 0 & -\lambda \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

El polinomio característico resulta ser $|A - \lambda I| = \lambda^4 - \frac{3\lambda^2}{4} - \frac{\lambda}{4} = \lambda(\lambda - 1)\left(\lambda + \frac{1}{2}\right)^2$.

Se observa que sus raíces son 0, 1 y $-\frac{1}{2}$ (esta última es una **raíz doble**), y que 1 es mayor que las otras raíces (en valor absoluto).

Si se resuelve el sistema para $\lambda = 1$, se puede ver que se trata de un **sistema uniparamétrico**, es decir, solo hay una **incógnita libre**; y, si se elige aquella solución cuyas componentes suman 1, se obtiene

$$\mathbf{v} = \begin{pmatrix} 2/9 \\ 1/3 \\ 5/18 \\ 1/6 \end{pmatrix} \approx \begin{pmatrix} 0.2222 \\ 0.3333 \\ 0.2778 \\ 0.1667 \end{pmatrix},$$

que ordena las páginas, de más a menos importante, así: 2ª, 3ª, 1ª, 4ª.

Ejercicio 6.2. Cuando tengas claros los conceptos y tengas herramientas adecuadas disponibles, realiza con detalle todos los cálculos de este ejemplo.

Discusión y trabajo a desarrollar

La teoría de grafos juega un papel clave en la moderna rama de Sistemas complejos. A su vez, las matrices y sus propiedades son esenciales en dicha teoría. En este artículo te he presentado dos ejemplos (no triviales), aunque de tamaño reducido. En el primer ejemplo, hemos visto una aplicación en clasificaciones deportivas. Aunque el ejemplo puede parecer poco útil, existen múltiples modelos semejantes en diversas ramas de la Ciencia. El segundo ejemplo, a pesar de ser 'de juguete' es la base del algoritmo PageRank que utiliza Google en las búsquedas por internet.

Tarea

1. Realiza con detalle todos los ejercicios propuestos.
2. Trata de generalizar (en términos de otras aplicaciones y, también, del tamaño de los problemas) los ejemplos presentados.
3. Quizá necesites escribir algún programa en Python para automatizar algunos de los cálculos.

7 Cierre

La teoría de grafos es esencial en el estudio de redes complejas. Las matrices son una herramienta clave en dicha teoría. El Álgebra proporciona las herramientas clave en la teoría de grafos. El contenido de este artículo ha sido impartido en una clase inicial en la asignatura Álgebra del Grado en Ingeniería Física de la Universitat Politècnica de València en el curso 2021/2022, proponiendo, a su vez, un trabajo académico con valor evaluativo, que ha producido excelentes resultados, como muestra la calidad de los trabajos presentados y el interés despertado.

8 Bibliografía

Izquierdo Sebastián, Joaquín, Torregrosa Sánchez, Juan Ramón, Álgebra y ecuaciones diferenciales. Tomos I y II. Universitat Politècnica de València, Valencia, 1997, SPUPV 97.669.

Johnson, http://nilesjohnson.net/