Clasificación por orden de importancia de la producción científica: Modelamiento de un algoritmo matemático utilizando Python

Consejo Nacional de Ciencia, Tecnología e Innovación – CONCYTEC

Dirección de Investigación y Estudios – DIE

Jhon Moisés Collantes Rios

Especialista en Estudios Econométricos

Índice general

Capítulo 1

Resumen

El estudio de Clasificación por orden de importancia de la producción científica: Modelamiento de un algoritmo matemático utilizando Python, nace de la necesidad de descifrar y medir algunas de las métricas de interés del documento "Principales indicadores bibliométricos de la actividad científica peruana. 2012-2017"(Moya, F. et. al, 2019), a fin de clasificar por orden de importancia la producción científica, ya sea por artículos, autores o revistas científicas a través del desarrollo de algoritmos matemáticos.

Por lo que, la fuente primaria de información (citas) al igual que el documento de Moya es recolectada de la base de datos de Scopus.

Por otro lado, la información que se requiera de otras instituciones dependerá de cuestiones como desempeño del algoritmo, los tiempos de respuesta del servidor, estructura de los datos recopilados, entre otros.

La naturaleza de los datos conlleva a la disyuntiva del método de investigación para el estudio. En principio los datos nos imponen variables nominales, por ejemplo, nombres de artículos, autores, etc., sin embargo, la parte importante del estudio tiene que ver con la referenciación que los involucrados hagan entre sí. Ante esto se propone una metodología cuantitativa antes que cualitativa, debido a la naturaleza del parámetro más importante del estudio, la referenciación, esta es una variable cuantitativa que por las interconexiones que puede haber, por ejemplo, entre artículos, se vuelve muy compleja de tratar.

En ese sentido, se ha modelado un algoritmo que se fundamenta en la teoría matemática de manera rigurosa utilizando el software libre **Python**, para así desarrollar y entender el algoritmo que permita clasificar por orden de importancia la producción científica tanto de un artículo, revista científica o autor.

Capítulo 2

Introducción

El documento "Principales indicadores bibliométricos de la actividad científica peruana. 2012-2017" (Moya, F. et. al, 2019), señala que la investigación científica desarrollada en Perú se caracteriza mediante la determinación del grado de visibilidad, colaboración, impacto, excelencia y liderazgo que alcanzaron los investigadores peruanos. Asimismo, los indicadores más utilizados para medir la producción científica se basan en los recuentos de las publicaciones y de las citas recibidas por los trabajos publicados, así como en el impacto de las revistas de publicación. El factor de impacto (FI) de las revistas es un indicador presente en casi todos los análisis de la producción científica; aunque en ocasiones se equipara -alto factor de impacto- con mayor calidad, este indicador mide específicamente la visibilidad y la difusión de los trabajos publicados en estas revistas más que la calidad científica de los mismos.

Si visualizamos todos los artículos, revistas o autores en una gran base de datos, ciertamente nos encontramos ante un conjunto estático y no muestra relevancia alguna para cualquiera sea la tarea que nos lleve a tal base de datos; este conjunto gana dinamismo y relevancia gracias a la referenciación entre sus elementos. Estamos ante *una (gran) red*.

Para traducir de manera objetiva esta relevancia requerimos de ramas del conocimiento como la **Teoría de Redes** que busca patrones concretos y medibles de relaciones entre entidades en un espacio, que generalmente es social. El parámetro que utilizaremos para clasificar un autor, artículo o revista será la cantidad y calidad de los autores, artículos y/o revistas que lo referencien. Así, la red que tenemos a nuestra disposición es recogida por la teoría de grafos dirigidos.

Cabe indicar que en un primer momento la teoría involucrada para resolver el problema fue el Análisis de Citas, pero esta tiene sus fundamentos en bases de datos relacionales que muestran desventajas al análisis que deseamos hacer, entre ellos el poco dinamismo y la dificultad de comprender los datos cuando esta aumenta su tamaño de manera considerable; según varios artículos consultados la herramienta idónea involucra una que fue diseñada para la Web, el algoritmo PageRank.

El algoritmo PageRank fue desarrollado por Larry Page y Serguéi Brin en 1996, cuando era doctorandos en Informática en la Universidad de Stanford (EEUU), y es usado por Google en el ranking de páginas webs de su buscador, aunque no es el único algoritmo usa-

do por Google para ordenar sus búsquedas, es el primer algoritmo usado por la compañía y también el más conocido. Indagando sobre la importancia de los algoritmos que hacen trabajar el buscador de Google, PageRank sigue teniendo una importancia fundamental a la hora de destacar en los resultados de búsquedas de Google, según Carreras Lario, R. (2012). Cómo clasifica Google los resultados de las búsquedas: factores de posicionamiento orgánico [Tesis doctoral, Universidad Complutense de Madrid].

El algoritmo está basado en un teorema que envuelve al Álgebra Lineal, sin embargo, el planteamiento mismo del algoritmo involucra la Teoría de grafos y Teoría de probabilidad.

En cuanto al Álgebra Lineal el resultado más importante que se desarrollará es El teorema de Perron-Frobenius, demostrado en 1912 por Ferdinand Georg Frobenius, que mostró (aseguraba bajo cierta hipótesis) la existencia de un objeto matemático llamado **autovector positivo** que no es otra cosa que un arreglo de una cantidad finita de números positivos, aquí yace la idea que aprovecharon los fundadores de Google para su algoritmo. Para un entendimiento del teorema se revisará teoría de matrices, autovalores, autovectores y un poco de teoría de diagonalización de matrices.

En cuanto a la Teoría de grafos, aquí recae parcialmente el trabajo para poder satisfacer la hipótesis del teorema en cuestión. Por un lado, requerimos de un objeto matemático llamado **matriz irreducible** con entradas no negativas (la hipótesis) y por otro tenemos una gran base de datos (artículos, autores, revistas, etc.) en la cual sus elementos se referencian unos con otros no exhaustivamente, la teoría de grafos nos permite transformar esta información plasmada en la base de datos y vista como un grafo dirigido (flechas que van de un elemento a otro y que siguen la dirección del que referencia) a un matriz, conocida como matriz de adyacencia.

En cuanto a la Teoría de probabilidad, aquí recae finalizar el trabajo para poder satisfacer la hipótesis. La matriz de adyacencia, que es un arreglo rectangular de dimensiones inmensas, generalmente no cumple la condición de ser irreducible (esta es una condición muy difícil de cumplir y más aún verificar) esto está íntimamente ligado a como se conectan los elementos (nodos) del grafo, de haber elementos que no se conecten entre sí (no se referencien) esto puede llevar a tener muchos ceros en la matriz y por ende tener una matriz que no es irreducible. Para evitar este inconveniente la teoría de probabilidad, gracias a la ayuda del llamado **vector de probabilidad**, modifica nuestra previa matriz de adyacencia, asegurando no tener muchos ceros en la matriz, por ende, el grafo se vuelve **más denso**, así aseguramos la hipótesis de tener una matriz irreducible con entradas no negativas.

Estas tres teorías matemáticas nos servirán para asegurar la existencia del autovector positivo o PageRank, hasta ahora el enfoque podría parecer más teórico que práctico; sin embargo, para la obtención de tal vector necesitaremos potentes métodos de cálculo que involucran convergencia de matrices. Para tal fin veremos El Método de las potencias y sus variantes.

En el apéndice A.1 y A.2 se demuestra el Teorema que justifica el trabajo, el Teorema de Perron - Frobenius.

Los últimos capítulos se avocan a poner a prueba el algoritmo, los resultados vendrán acompañado de sus interpretaciones. Las conclusiones irán acordes a lo investigado y las recomendaciones serán direccionadas a mejoras para el algoritmo, así como también considerar otros parámetros, dependiendo del desempeño del algoritmo, así como su adecuación.

Capítulo 3

Estado del Arte de la Investigación

En La Historia de la Ciencia ocurre un tipo de suceso común en las ciencias básicas, este es el de desarrollar conocimiento teórico que en su descubrimiento o invención no tiene aplicación práctica alguna, y por décadas o incluso siglos pasan "durmiendo el sueño de los justos" hasta que mentes disidentes en pos del avance científico o tecnológico finalmente los utilizan.

Esto no es ajeno al principal resultado que el estudio utilizará para sus propósitos, el Teorema de Perrón-Frobenius, "a lo largo de los últimos 100 años, se ha demostrado cómo un resultado esencialmente teórico que puede tener potentes aplicaciones a muy diferentes ámbitos de la ciencia y la tecnología." (R. Criado, M. Romance y L. E. Solá, 2014). Así pues, estamos ante una poderosa herramienta que para su enunciación tiene ideas muy simples, pero para su demostración tuvo que apoyarse de otros resultados, descubierto por Luitzen Bouwer, que desarrollaremos más adelante en este informe, "Uno de los ingredientes fundamentales de la prueba del Teorema de Perron-Frobenius es el Teorema del punto fijo de Brouwer. Este teorema, que data del 1910, nos asegura la existencia de un punto fijo para una aplicación continua definida en un conjunto compacto. La idea del teorema le vino ingerida a Brouwer cuando removía una taza de café, en la que aparentemente siempre había un punto sin movimiento. Este teorema de apariencia simple, pero de difícil demostración, ha sido la pieza fundamental de la prueba de diferentes resultados, a parte del Teorema de Perron-Frobenius." (Inés A. 2015). El teorema de Perron-Frobenius y su aplicación en el algoritmo de búsqueda de Google).

Los métodos de demostración de este teorema auxiliar son variados por ejemplo puede estar basado en el juego de Hex, inventado por John Nash, también podemos usar topología combinatoria, o resultados de la geometría diferencial como el Teorema de Stokes o el Teorema de la bola peluda. Un aporte a estas demostraciones es recogido en el trabajo de Edson Quispe Calderón, en su trabajo "Prueba geométrica del Teorema de Perron-Frobenius y su aplicación a Google, Universidad Nacional del Altiplano, 2014".

Respecto al modelo prototipo, PageRank, tienes sus orígenes en el pionero artículo de los fundadores de Google titulado "The anatomy of a large-scale hypertextual Web search engine, Stanford University, 1998", la siguiente cita de este artículo sigue siendo relevante hasta nuestros días: "*La medida más importante de un motor de búsqueda es la calidad de sus resultados de búsqueda. Si bien una evaluación completa del usuario está más allá*

del alcance de este documento, nuestra propia experiencia con Google ha demostrado que produce mejores resultados que los principales motores de búsqueda comerciales para la mayoría de las búsquedas. Como ejemplo que ilustra el uso de PageRank", p. 113.

La mayoría de fuentes consultadas vuelven en gran medida a este artículo, ciertamente cuando uno se familiariza con las "características del modelo" puede dar fe de lo siguiente: "... es un tema muy discutido por los expertos en optimización de motores de búsqueda (SEO, por sus siglas en ingles). En el corazón de PageRank hay una fórmula matemática que parece aterradora de ver, pero que en realidad es bastante simple de entender." (I. Rogers, The Google Pagerank Algorithm and How It Works).

Lo que ha ido cambiando y especializando son distintos campos en los que se puede aplicar este algoritmo, desde rankings en la Liga Nacional de Fútbol Americano ("An application of Google PageRank to NFL rankings, Involve a journal of mathematics, L. Zack, R. Lamb, S. Ball, 2012") hasta estudios sobre el flujo del tránsito (B. Jiang, S. Zhao, J. Yin, "Self-organized Natural Roads for Predicting Traffic Flow: A Sensitivity Study, The Hong Kong Polytechnic University, 2008"). Los artículos consultados que muestran parte de su desarrollo computacional tienen variaciones entre sí. Esto muestra la divergencia de los métodos computacionales de acuerdo al campo en que se desarrollan.

Dos trabajos relacionados con el tercer entregable son el de Laura Miralles Millas en "Teorema de Perron-Frobenius Aplicación en el JCR (Eigenfactor TM Score en el JCR), Universidad de Zaragoza 2019", aquí toma relevancia un indicador bibliométrico llamado Eigenfactor Score (EFS) en comparación otro más tradicional, el Factor de Impacto (FI). Un uso de ambos indicadores se utiliza en diversos rankings de revistas en el Journal Citation Reports (JCR).

El segundo es S. Maslov, S. Render en "Promise and Pitfalls of Extending Google PageRank" del Departamento de Física de la Materia Condensada y Ciencias de los materiales, New York; este artículo menciona las ventajas y advierte de los peligros que tienen el usar indiscriminadamente el algoritmo PageRank en la referenciación de revistas y artículos. Comparado con el artículo anterior, aquí hay análisis de cientos de miles de artículos, pero debe entenderse más como un informe comparativo que un trabajo donde se puedan aprovechar técnicas para los cálculos que se necesiten.

Escoger el método computacional de nuestro modelo es parte esencial del cuarto entregable por lo que las fuentes recolectadas hasta el momento se tornan prometedoras.

Resulta de gran interés los últimos avances del algoritmo PageRank del siguiente artículo de S. Park, W. Lee, B. Choe, S. Lee titulado "A Survey on Personalized PageRank Computation Algorithms, 2019" de la Universidad Nacional de Seúl, Corea del Sur; donde además se muestran otros métodos computacionales, siendo el más común y al que la mayoría de los artículos anteriores se apoyaban el Método de las potencias (señalado en aquí como Iterative Equation). Inclusive se muestra una comparación completa de los nuevos métodos computaciones.

	Precomputation Time
Iterative Equation Solving	None
Direct Equation Solving	Very High
Bookmark Coloring	None
Dynamic Programming	Very High
Monte-Carlo Sampling	High

Figura 3.1: Summary of Personalized PageRank computation approaches.

Capítulo 4

Planteamiento del modelo

Supongamos que tenemos n objetos, que pueden ser revistas, artículos o autores, denotados por P_i, $i = 1, \ldots, n$ y por $x_i(P_i)$ la importancia del artículo P_i. Es claro que en una comunidad científica estos objetos deben relacionarse entre ellos, citándose entre ellos, por ejemplo, un artículo cita el trabajo de otro dos artículos, o haciendo referencia uno del otro en bibliografía final del documento científico.

Se entenderá por referenciación a la acción de haber sido listado como fuente consultada para la elaboración de uno o varios trabajos de índole científico.

Para este trabajo se entiende que las revistas se referencian entre ellas, los artículos se referencian entre ellos, los autores se referencian entre ellos; la referenciación entre estos tres tipos de objetos tiene sus propias características de acuerdo a su naturaleza, que se hará mención en la siguiente sección, pero esto no repercute en la estructura del modelo que se muestra a continuación. Se verá un caso particular.

Para este fin se supone que tenemos n artículos denotados por P_i, $i = 1, \ldots, n$ y por x_i la importancia del artículo P_i. Teniendo en cuenta la referenciación (que también llamaremos enlaces) entre artículos, se puede considerar estas relaciones como un grafo dirigido (P, E), cuyos vértices serán los artículos y las aristas serán la referenciación entre ellas, el artículo P_3 referencia al artículo P_5, esto se traduce a una arista dirigida.

Tenemos un conjunto de artículos $P_i : \{P_1, \ldots, P_n\}$, y un vector de importancias:

$$\begin{bmatrix} x_1(P_1) \\ x_2(P_2) \\ \vdots \\ x_n(P_n) \end{bmatrix}$$

Para llevar el grafo a un objeto apto para cálculos utilizamos su matriz de adyacencia, ordenamos los vértices $P_1, \ldots, P_n$ y le asignamos a cada uno de ellos una fila y una columna en una matriz cuadrada, la llamaremos primera matriz de adyacencia del modelo. Se obtiene:

$$M = \begin{array}{c} \\ P_1 \\ \vdots \\ P_i \\ \vdots \\ P_n \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & \dots & P_j & \dots & P_n \end{array} \\ \begin{pmatrix} \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & \dots & m_{ij} & \dots & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & \dots & \vdots & \dots & \vdots \end{pmatrix} \end{array}$$

con la siguiente regla de correspondencia para llenar las entradas de la matriz

$$m_{ij} = \begin{cases} 1 & \text{, si hay enlace de } P_j \text{ a } P_i \\ 0 & \text{, si no hay enlace de } P_j \text{ a } P_i \end{cases}$$

Para saber el número de enlaces que hay a cada artículo, servirá con sumar las entradas de la fila correspondiente. De manera análoga la columna i-ésima representa los enlaces que tiene el artículo P_i a los diferentes artículos.

Tener en cuenta que la matriz M que vamos a utilizar no es simétrica.

Para facilitar el planteamiento del modelo se consideran un conjunto formado por cinco artículos P_i, $i = 1, \dots, 5$, que se relacionan como muestra la figura (4.1), esta información se puede plasmar en un grafo de la figura (4.2)

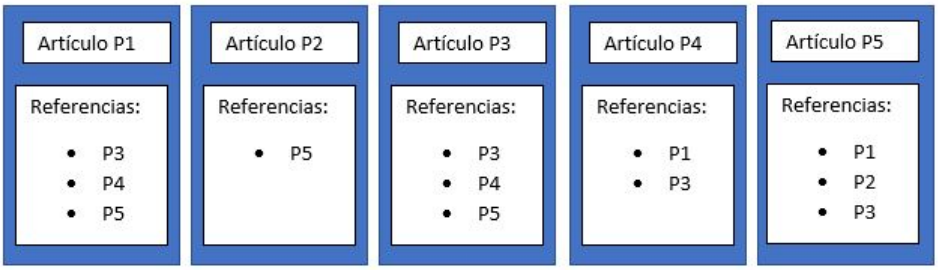

Figura 4.1: Relación entre artículos

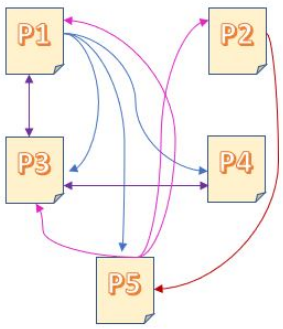

Figura 4.2: Grafo asociado a los artículos

Y tendrá la siguiente matriz de adyacencia:

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Sin embargo, para cuestiones computacionales, es mejor tener una matriz estocástica para eso vamos a modificar M. Esta es la matriz base del modelo.

$$M^* = \begin{array}{c} \\ P_1 \\ \vdots \\ P_i \\ \vdots \\ P_n \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & \cdots & P_j & \cdots & P_n \end{array} \\ \begin{pmatrix} \vdots & \cdots & \vdots & \cdots & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & \cdots & \frac{m_{ij}}{n_j} & \cdots & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & \cdots & \vdots & \cdots & \vdots \end{pmatrix} \end{array}$$

donde m_{ij} es como se definió anteriormente y n_{ij} es el número de enlaces al artículo P_j.

Entonces para el caso M la nueva matriz que se obtiene es:

$$M^* = \begin{bmatrix} 0 & 0 & 1/3 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 1 & 1/3 & 0 & 0 \end{bmatrix}$$

este un ejemplo de la matriz base del modelo para un caso de 5 artículos.

Ahora el planteamiento del modelo, es decir, cuantificar las importancias x_i. La manera más fácil, e incorrecta, de plantear esto sería simplemente sumar la cantidad de enlaces, por ejemplo, el grafo mostrado tendría el vector de importancias $(3, 1, 2, 2, 3)$. Se puede percibir esta no es la situación que se pretende, ya que puede haber páginas que son referenciadas desde páginas muy importantes, pero que tengan pocas referencias.

Lo que se hará consiste en considerar la importancia x_i de cada artículo P_i como directamente proporcional a la suma de las importancias de los artículos que la referencian. De esta forma está claro que evitamos el problema que mostrábamos anteriormente.

En esta situación para los artículos de ejemplo se tendría que las importancias seguirán las siguientes relaciones:

$$\begin{cases} x_1(P_1) = \frac{1}{\alpha}(x_3(P_3) + x_4(P_4) + x_5(P_5)) \\ x_2(P_2) = \frac{1}{\alpha}(x_5(P_5)) \\ x_3(P_3) = \frac{1}{\alpha}(x_1(P_1) + x_4(P_4)) \\ x_4(P_4) = \frac{1}{\alpha}(x_1(P_1) + x_3(P_3)) \\ x_5(P_5) = \frac{1}{\alpha}(x_1(P_1) + x_2(P_2) + x_3(P_3) \end{cases}$$

donde $\frac{1}{\alpha}$ es cierta constante de proporcionalidad. Que representado en forma matricial tendrá el siguiente aspecto.

$$\begin{bmatrix} x_1(P_1) \\ x_2(P_2) \\ x_3(P_3) \\ x_4(P_4) \\ x_5(P_5) \end{bmatrix} = \frac{1}{\alpha} \begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(P_1) \\ x_2(P_2) \\ x_3(P_3) \\ x_4(P_4) \\ x_5(P_5) \end{bmatrix}$$

Si se denota por x el vector que contiene las importancias de los artículos y llamamos λ a $\frac{1}{\alpha}$ podremos escribir la asignación de importancias que pretendemos obtener en una solución de la ecuación:

$$M^* x = \lambda x$$

que bajo ciertas manipulaciones se representa por:

$$(M^* - \lambda I)\, x = 0$$

donde I es la matriz identidad de orden n.

La cuestión se ha transformado en un problema de autovalores y autovectores, de esta forma llegamos a que el vector de importancias buscado, es un autovector de la matriz M^* que por supuesto debe de tener todas sus entradas positivas.

En estos momentos es cuando nos damos cuenta de que el teorema de Perron-Frobenius que hemos tratado anteriormente, puede tener una gran importancia, ya que bajo ciertas condiciones nos asegura la existencia del vector buscado.

Una vez visto que nos encontramos en una situación en la que puede que nos sea útil el teorema de Perron-Frobenius, es hora de comprobar que realmente podemos aplicarlo, es decir que estamos ante una matriz no negativa e irreducible.

A eso se avocan los esfuerzos en la siguiente sección.

Capítulo 5

Desarrollo del modelo

Tenemos 3 tipos de objetos, revistas, artículos y autores; como se dijo en la sección anterior, esto no afecta la estructura del modelo, pero si se deben hacer ciertas observaciones. Primero de tipo general:

1. El periodo que considera como referencia para contarse como tal es de cinco años. Es un período de tiempo que usan muchos rankings de tipo bibliográfico.
2. No se tiene en cuenta las autorreferencias, esto sobre todo en el caso de las revistas. Así se evita que los objetos involucrados inflen sus categorías o que objetos poco usuales se comporten como un grafo aislado del resto.
3. El modelo pondera la importancia de las referenciaciones recibidas por un objeto por la importancia de los objetos que lo referencian.
4. Por distintos que sean los números de las entradas de la primera matriz de adyacencia, la matriz base del modelo es estocástica, por lo que sus entradas serán menores que uno.

Ahora observaciones de tipo particular:

- Para las revistas:
  - Al ser un cúmulo de artículos la negativa de contar autorreferencias podría jugar un papel preponderante en el ranking, esto sin embargo no sucede.
  - Las entradas de la primera matriz de adyacencia del modelo, en general, tiene entradas positivas distintas de 1 ya que la referenciación entre revistas implica la referenciación entre los artículos de la revista correspondiente y estos son más de uno por revista.
- Para los artículos:
  - Las entradas de la primera matriz de adyacencia del modelo, en general, tiene entradas 1 y 0 ya que la referenciación entre artículos es de índole individual. Esto es de gran ayuda cuando queremos ejemplificar el desarrollo del modelo.
- Para los autores:
  - Las entradas de la primera matriz de adyacencia del modelo, en general, tiene entradas positivas distintas de 1, a pesar que el objeto autor se considere de índole individual, un autor puede referenciar a otro en diferentes ocasiones.

Por la sencillez de la primera matriz de adyacencia del objeto artículo continuamos desarrollando el modelo para tal caso.

Ya se ha modelizado una matriz de adyacencia relacionado con artículos y sus referenciaciones,*sin embargo, el desarrollo del modelo implica poder asegurar la existencia de un vector que funja de ordenamiento*, es decir, la matriz que utilicemos cumpla las hipótesis del Teorema de Perron-Frobenius y así asegurar la existencia del autovector de coordenadas positivas relacionado con el autovalor de mayor tamaño.

Para el ejemplo previo, la matriz ya muestra dos bloques de tamaño 2×2 con entradas cero,

$$M^* = \begin{bmatrix} 0 & 0 & 1/3 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 1 & 1/3 & 0 & 0 \end{bmatrix}$$

por lo que bastaría encontrar cierta permutación y notar que la matriz A es reducible, cosa que no es deseable. Sin embargo, obtener una matriz de ese tipo es algo normal, ya que artículos de temas especializados o pioneros en los tópicos que traten tendrán pocas referenciaciones, dependiendo del tema, fecha de publicación o búsqueda cuando utilicemos nuestro algoritmo (si hacemos un ranking de cierto tipo de artículos de un artículo brillante, pero tiene poco tiempo de haberse publicado quizás aún no haya sido referenciado debidamente por su comunidad o círculo científico).

Se modifica un poco más el modelo creando una matriz M'' de la forma:

$$M'' = \alpha M' + (1 - \alpha)\, ve^t \tag{5.1}$$

donde α es un valor entre 0 y 1, $e^t = (1, \ldots, 1)$ y v un vector de probabilidad que suele ser definido como $v = \frac{1}{n}e$.

Así se consigue una matriz M'' irreducible en todos los casos a costa de hacerla densa[1] . El término introducido $(1 - \alpha)\, ve^t$ se puede considerar como la probabilidad de que un investigador decida ya no seguir las referencias de su artículo y acuda a uno totalmente aleatorio. Es claro que cuanto más cercano a 1 sea el valor de α, más nos acercaremos a la matriz del grafo de original, pero a costa de poder perder la irreducibilidad.

En la sección anterior, para el ejemplo de 5 artículos, se mostró una matriz base para el modelo, M^*, pero tiene el inconveniente de ser reducible, para evitar eso se modifica usando la expresión (5.1), el procedimiento y la forma matricial tiene el siguiente aspecto.

[1]La densidad es una propiedad topológica que en el argot matemático puede ser entendida como la noción de no tener zonas vacías, en nuestro contexto dar densidad es evitar tener bloques de entradas nulas en las matrices

$$M'' = 0{,}85 \begin{bmatrix} 0 & 0 & 1/3 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 1 & 1/3 & 0 & 0 \end{bmatrix} + (1 - 0{,}85)(1/5) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}^t$$

Desarrollando las operaciones se obtiene:

$$\begin{bmatrix} 0{,}0300 & 0{,}0300 & 0{,}3133 & 0{,}4550 & 0{,}4550 \\ 0{,}0300 & 0{,}0300 & 0{,}0300 & 0{,}0300 & 0{,}4550 \\ 0{,}3133 & 0{,}0300 & 0{,}0300 & 0{,}4550 & 0{,}0300 \\ 0{,}3133 & 0{,}0300 & 0{,}3133 & 0{,}0300 & 0{,}0300 \\ 0{,}3133 & 0{,}8800 & 0{,}3133 & 0{,}0300 & 0{,}0300 \end{bmatrix}$$

Construir de esta forma la matriz se consigue que la matriz siempre sea irreducible, así el teorema de Perron-Frobenius asegura la existencia del vector de ordenación.

El inconveniente actual es que la demostración del teorema se consigue bajo un proceso no constructivo, por lo que se debe acudir a métodos auxiliares para obtener el deseado vector, además nuestros cálculos involucran matrices de dimensiones muchísimo más grandes que 5, es decir, la búsqueda de los autovectores no puede ser exhaustiva.

Sin embargo, como aún se está desarrollando el modelo se usa como ejemplo una matriz M'' de dimensión baja.

Capítulo 6

Cálculo del ordenamiento

Para calcular el vector de ordenación se hará uso del Método de las Potencias. Se ejemplificará con un conjunto de cuatro artículos. *Se ordenarán los siguientes cuatro artículos*:

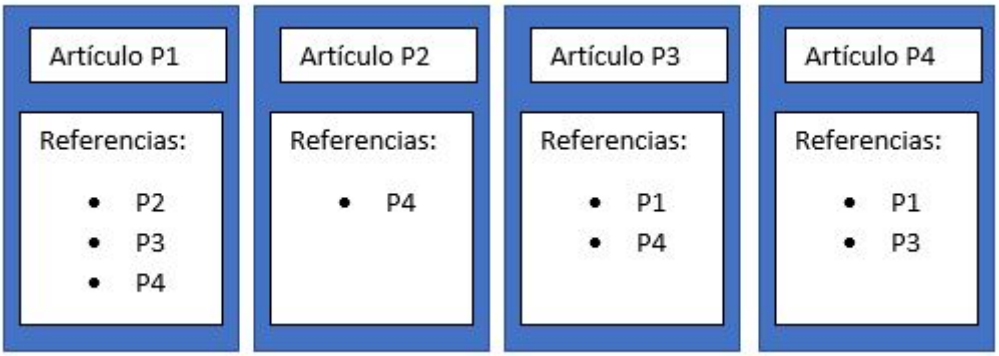

Figura 6.1: Relación entre 4 artículos

La matriz asociada es:

$$M = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Claramente aquí se presenta un bloque, nuevamente, de orden 2 que hace a M reducible, arreglamos esto, pero aún llamamos M a la matriz:

$$M = 0{,}85 \begin{bmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{bmatrix} + (1 - 0{,}85)(1/4) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}^t$$

Como el método de las potencias consiste en iterar un vector inicial x_0, se define ese vector como la norma de Frobenius de la matriz M. Para eso realizamos este producto:

$$M^t M = \begin{bmatrix} 0{,}3102 & 0{,}3102 & 0{,}1898 & 0{,}1898 \\ 0{,}3102 & 0{,}7919 & 0{,}4306 & 0{,}0694 \\ 0{,}1898 & 0{,}4306 & 0{,}4306 & 0{,}2500 \\ 0{,}1898 & 0{,}0694 & 0{,}2500 & 0{,}4306 \end{bmatrix}$$

Nos interesa la diagonal de esta matriz:

$$h_1 = \begin{bmatrix} 0{,}3102 \\ 0{,}7919 \\ 0{,}4306 \\ 0{,}4306 \end{bmatrix}$$

Sin embargo, por conveniencia, normalizamos, es decir, dividimos entre su norma:

$$x_0 = \frac{1}{1{,}0460} \begin{bmatrix} 0{,}3102 \\ 0{,}7919 \\ 0{,}4306 \\ 0{,}4306 \end{bmatrix} = \begin{bmatrix} 0{,}2966 \\ 0{,}7570 \\ 0{,}4117 \\ 0{,}4117 \end{bmatrix}$$

Esta es la aproximación inicial del vector de importancias que se busca.

La siguiente aproximación viene dada por la ecuación:

$$x_1 = \frac{\left(M^t M\right) x_0}{\| \left(M^t M\right) x_0 \|}$$

Desarrollando, se obtiene:

$$x_1 = \frac{1}{1{,}2762} \begin{bmatrix} 0{,}4831 \\ 0{,}8973 \\ 0{,}6625 \\ 0{,}3890 \end{bmatrix} = \begin{bmatrix} 0{,}3785 \\ 0{,}7031 \\ 0{,}5191 \\ 0{,}3048 \end{bmatrix}$$

Las siguientes aproximaciones se calculan con la siguiente fórmula, dando paso a un proceso iterativo:

$$x_i = \frac{\left(M^t M\right) x_{i-1}}{\| \left(M^t M\right) x_{i-1} \|}$$

Para una aproximación inicial al autovalor máximo (λ), multiplicamos el vector x_1 y su vector antecesor x_0:

$$\lambda^1 = x_1^t x_0 = \begin{bmatrix} 0{,}3785 \\ 0{,}7031 \\ 0{,}5191 \\ 0{,}3048 \end{bmatrix}^t \begin{bmatrix} 0{,}2966 \\ 0{,}7570 \\ 0{,}4117 \\ 0{,}4117 \end{bmatrix} = 0{,}9837$$

De forma similar se calcula $\lambda^2, \ldots, \lambda^i$ para aproximar el autovalor máximo con la fórmula:

$$\lambda^{(i)} = x_i^t x_{i-1}$$

Calculando la iteración número dos se obtiene:

$$x_2 = \begin{bmatrix} 0{,}3788 \\ 0{,}7075 \\ 0{,}5192 \\ 0{,}2939 \end{bmatrix}$$

$$\lambda^2 = x_2^t x_1 = \begin{bmatrix} 0{,}3788 \\ 0{,}7075 \\ 0{,}5192 \\ 0{,}2939 \end{bmatrix}^t \begin{bmatrix} 0{,}3785 \\ 0{,}7031 \\ 0{,}5191 \\ 0{,}3048 \end{bmatrix} = 0{,}9999$$

Se tiene una sucesión de iteraciones, que en algún momento debería de parar, para ello se debe de considerar que el error relativo estimado sea menor al valor de tolerancia asignado (E), por lo general es un número decimal que se aproxime a cero hasta en 16 casas decimales (10^{-16}):

$$\left| \frac{\lambda^{(n)} - \lambda^{(n-1)}}{\lambda^{(n)}} \right| < E$$

donde

$\lambda^{(n)}$ = Aproximación al autovalor en la iteración n
$\lambda^{(n)}$ = Aproximación al autovalor en la iteración $n-1$
E = Valor de tolerancia

En el apéndice se procede a calcular la secuencia, esta terminará cuando el error relativo sea menor que E, muestra el código en Python y con fines didácticos señalar lo siguiente:

- Número de iteración
- Vector de importancias
- Error relativo en cada iteración

En la iteración diecisiete el error relativo estimado es menor al valor de tolerancia, se detiene la sucesión.

El autovector de importancia es:

$$v = \begin{pmatrix} 0{,}37791501 \\ 0{,}71084549 \\ 0{,}51824671 \\ 0{,}28861614 \end{pmatrix}$$

Por lo que los artículos pueden ser listados de la siguiente manera: P_2, P_3, P_1, P_4.

Capítulo 7

Clasificación por importancia de la producción científica

7.1. Recojo de datos

Los datos para las siguientes clasificaciones son recabados del portal de la editorial académica SCOPUS, una base de datos bibliográficos de resúmenes y citas de artículos de revistas científicas en línea, cubre aproximadamente 24.500 títulos de publicaciones seriadas (revistas, conferencias, series de libros de investigación) de más de 5000 editores en 140 países, incluyendo revistas revisadas por pares de las áreas de ciencias, tecnología, medicina y ciencias sociales, incluyendo artes y humanidades.

Una herramienta de evaluación del desempeño investigador que permite distintos tipos de análisis es SciVal. Toma los datos de la base de datos de SCOPUS, que cubre más de 78 millones de registros de trabajos científicos publicados desde 1996, de más de 24.000 revistas de más de 5.000 editores de todo el mundo.

SciVal da acceso al desempeño investigador de instituciones investigadoras, así como investigadores individuales de todo el mundo y permite:

1. Visualizar dicho desempeño.
2. Establecer diversas comparativas a través de distintas métricas.
3. Revisar redes de coautoría.
4. Identificar potenciales asociaciones de colaboración científica.
5. Ayudar a la planificación de la investigación a nivel de centros y departamentos

7.2. Clasificación de artículos

Vamos a enfocarnos en el siguiente nicho de publicaciones, se aplicará el siguiente filtro, y vamos a realizar un ranking por importancia:

A. Tópico: Medicina

B. Recurso SCOPUS: Revista Peruana de Medicina Experimental y Salud Pública

Data set	Publications in Peru
year range	2017 to 2021
Subject classification	ASJC
Filtered by	Medicine
Types of publications included	All publication types
Self-citations	–
Data source	Scopus
Date last updated	7 September 2022
Date exported	15 September 2022
373 publications match the selected filter options:	
Countries/Regions	Peru
Publications types	Article
Scopus Sources	Revista Peruana de Medicina Experimental y Salud Pública

Cuadro 7.1: Información del filtro proporcionada por SCOPUS.

C. Tipo de Publicación: Artículo

D. Rango de tiempo: 2017 - 2021, últimos 5 años.

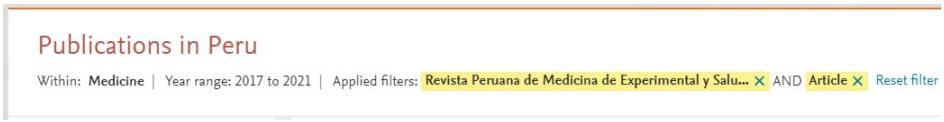

Figura 7.1: Primer ejemplo para analizar, ver filtros aplicados. Web de SCOPUS.

El filtro arroja un total de 373 artículos, vamos a listar mostrar los cuatro primeros resultados:

1. Mental health considerations about the COVID-19 pandemic
2. Tuberculosis in Peru: Epidemiological situation, progress and challenges for its control
3. Scientific production and licensing of medical schools in Peru
4. Preliminary results of the strengthening of the national death registry information system

La extracción de la información que se utilizará para clasificar artículos, revistas y autores se puede realizar de manera manual o con las APIs de Scopus, esta última herramienta requiere de un registro en su Web y para usar la mayoría de funcionalidades de las APIs se requiere de una suscripción institucional, ya que, consumir la información de Scopus mediante la API requiere de una clave institucional (Insttoken).

Para el primer caso, la web de Scopus permite exportar la lista de artículos a un archivo .xlsx, la información detallada proporcionada por SCOPUS del filtro aplicado se puede ver en el cuadro (7.1).

Vamos a crear la matriz de adyacencia, Definición (A.1.4), inicial con la relación que nos muestre SCOPUS sobre los artículos involucrados.

Al hacer clic en *View in SCOPUS*, ver figura (7.2), recogemos información de los siguientes hipervínculos.

- Cited by
- References

Title	Authors	Year	Scopus Source	Citations
Mental health considerations about the COVID-19 pandemic Open Access View abstract View in Scopus		2020	Revista Peruana de Medicina Experimental y Salud Publica	74
Tuberculosis in Peru: Epidemiological situation, progress and challenges for its control Open Access View abstract View in Scopus		2017	Revista Peruana de Medicina Experimental y Salud Publica	41
Scientific production and licensing of medical schools in Peru Open Access View abstract View in Scopus		2019	Revista Peruana de Medicina Experimental y Salud Publica	31
Preliminary results of the strengthening of the national death registry information system Open Access View abstract View in Scopus		2018	Revista Peruana de Medicina Experimental y Salud Publica	21

Figura 7.2: Lista de 4 primeros artículos de 373 en total. Web de Scopus.

Y podemos empezar a construir la matriz de adyacencia (Definición A.1.4), ver figura (7.3).

	Title	Authors		1	2	3	4	5	6
1	Mental health considerations about the COVID-19 par		1	0	0	0	0	0	0
2	Tuberculosis in Peru: Epidemiological situation, progr		2	0	0	0	0	0	0
3	Scientific production and licensing of medical schools		3	0	0	0	0	0	0
4	Preliminary results of the strengthening of the nation		4	0	0	0	0	0	0
5	Description of patients with severe COVID-19 treated		5	0	0	0	0	0	0
6	Effect of anemia on child development: Long-term co		6	0	0	0	0	0	0
7	Barriers to effective care in the referral hospitals of P		7	0	0	0	0	0	0
8	Type 2 diabetes mellitus in Peru: A systematic review		8	0	0	0	0	0	0
9	Evaluation under field conditions of a rapid test for de		9	0	0	0	0	0	0
10	Social programs and reducing obesity in Peru: Reflect		10	0	0	0	0	0	0

Figura 7.3: Creación de matriz base del modelo.

Al recabar la información de la base de datos guardamos la matriz sin encabezados ni columna vacía inicial en un archivo Excel con formato .csv y algún nombre conveniente, ya que utilizaremos esta matriz en el código del algoritmo. Para este caso la matriz cuadrada es de orden 373.

Para el segundo caso, utilizando las APIs de Scopus.

De acuerdo al Anexo (A.3), la información proporcionada por la API es por defecto un archivo en formato JSON, estos pueden ser fácilmente exportados a un archivo Excel, que tiene la ventaja de tener una interface más amena en la cual filtrar la sobreinformación recibida de SCOPUS. De esta manera se obtiene, como en el caso manual, el arreglo matricial (la matriz de adyacencia) requerido por el algoritmo.

El código encargado del ordenamiento se encuentra en el apéndice (A.5), sin embargo mostramos el vector de ordenamiento final, luego del truncamiento:

Listing 7.1: Primer Algoritmo de Prueba

```
Se necesitaron 195 iteraciones
Vector de ordenamiento de menor a mayor:
[[370 369 368 367   1 237 236 234 233 232 231 230 229 226 227 238 225 224
  223 222 221 220 228 239 241 219 267 266 265 264 263 259 258 257 256 255
  254 252 250 245 244 243 242 240 218 216 268 191 190 189 188 372 186 185
  184 182 181 180 179 178 177 176 175 174 192 217 194 197 215 214 213 212
  211 210 209 208 207 205 204 203 202 201 200 199 198 196 269 272 173 342
  341 339 338 337 336 335 334 333 332 331 329 327 326 325 323 322 343 321
  344 347 371 366 365 364 363 362 361 359 357 356 355 354 353 352 351 350
  348 345 271 320 317 292 291 290 288 287 286 285 283 282 281 279 278 277
  276 275 274 273 293 319 295 297 316 315 314 313 312 311 310 309 307 306
  305 303 302 301 300 299 298 296 172 187 170  74  75  24  76  23  22  79
   80  21  82  83 171  86  88  20  90  91 102  13 101  32 100  99  73  16
   17  95  18  19  94  93  97  72  25  71  50  49  48  47  46  45  52  29
   42  40  39  31  38  37  43 103  53  55  70  69  68  26  67  66  54  65
   62  27  60  59  57  56  63 104  33 106 150 149   7 147 146 144 143 142
  141 105 139 138 137 136 135 151   6 152 153   2 168 167   3 165   4 163
  134 162   5 159 158 157 156 155 154 161 133 140 120 115 132 118 119 114
  108 113  10 121 122 112 117 123 111 110 109 125 126 107 127   9 129 130
    8 124 116  36  28  15  12  87  92 247 261 145  41 340 160 360  64 195
  280 330 206 308 270  51 246  84 318  30 373 253 128 131  44 289  78  81
   14  34  58  89 164 294 284 183 249  98 166 235 148  85  77  11 358 169
   96 304 346  61 260 193 324 328  35 251 262 248 349]]
```

7.3. Clasificación de revistas

A partir de ahora seguimos los pasos de la clasificación anterior, así como el recojo de la información. En este caso la referenciación entre revistas es mayor, ya que estos involucran artículos por lo que comparado al caso anterior se observa lo siguiente:

1. La dimensión de la matriz es menor.

2. Las entradas de la matriz de adyacencia, en general, son mayores a 1, y la matriz puede ser, incluso, irreducible sin trabajo alguno.

En la siguiente clasificación nos atenemos a un nicho más específico, pudiendo clasificar todas las revistas sobre Matemáticas que nos pueda proporcionar nuestra base de datos, vamos a clasificar revistas de Matemática relacionadas al Álgebra. Los motivos para esto es notar lo siguiente: algunas métricas se ven afectadas por el tamaño de la entidad a clasificar, la disciplina, el tipo de publicación y el tiempo. Para nuestro algoritmo, la disciplina es quizás el factor más preponderante. La decisión de clasificar todas las revistas de Matemática puede ser un paso natural para todo usuario, sin embargo, nuestro algoritmo tiene su fortaleza en clasificar de acuerdo a como se relacionan los objetos involucrados; la Matemática y otras disciplinas tienen muchas subdisciplinas, las cuales se relacionan entre sí de manera muy diversa. Si están muy relacionadas esto puede ser beneficioso para ambas, o pueden ser tan especializadas y poco relacionadas que desde una visión general (todas las disciplinas) puede ser perjudicial para ellas en una clasificación.

Clasificar revistas de Matemática involucra tomar en cuenta subdisciplinas como Álgebra, Geometría, Teoría de números, Probabilidades, etc. Uno puede avocarse a clasificar todas o las de subdisciplinas en específico. Tal decisión queda a discreción del usuario luego de las observaciones mencionadas.

Para más detalle de los factores que afectan los valores de métricas se puede consultar [14].

En nuestro caso vamos a clasificar revistas españolas de Matemática relacionadas al Álgebra, en un período también de 5 años, desde el 2017, el resultado de tal filtro es el siguiente.

1. Journal of Algebra (J ALGEBRA)
2. Advances in Mathematics (ADV MATH)
3. Journal of pure and Applied Algebra (J PURE APPL ALGEBRA)
4. Algebras and Representation Theory (ALGEBRA REPRESENT TH)
5. Communications in Algebra (COMMUN ALGEBRA)
6. Journal of Group Theory (J GROUP THEORY)
7. Linear and Multilinear Algebra (LINEAR MULTILINEAR A)
8. Algebra Colloquium (ALGEBR COLLOQ)
9. Linear Algebra and its Applications (LINEAR ALGEBRA APPL)
10. Compositio Mathematica (COMPOS MATH)
11. Annals of Mathematics (ANN MATH)
12. Algebra and Number Theory (ALGEBR NUMBER THEORY)
13. Journal of Algebra and its Applications (J ALGEBRA APPL)
14. Journal of Commutative Algebra (J COMMUT ALGEBR)

0	49	43	34	134	22	10	8	18	6	2	10	74	9	0
73	0	18	10	20	2	3	7	13	13	3	9	3	1	16
93	22	0	8	38	2	0	1	0	0	0	0	25	4	5
44	6	3	0	39	0	9	0	0	0	0	0	17	0	0
39	10	13	8	0	7	14	11	112	0	0	0	88	6	0
11	1	2	4	10	0	0	4	0	0	0	0	6	0	1
9	0	0	0	20	0	0	9	76	0	0	0	0	0	0
7	0	0	0	20	0	0	0	0	0	0	0	10	0	0
0	0	0	0	55	0	174	10	0	0	0	0	25	0	0
16	38	4	3	9	0	0	0	0	0	4	17	0	0	0
15	48	7	1	4	5	0	1	0	20	0	11	0	0	7
18	12	2	5	2	0	0	0	0	16	0	0	0	0	3
26	0	10	6	64	5	5	4	6	0	0	0	0	0	0
3	0	0	0	4	0	0	2	0	0	0	0	7	0	0
4	22	3	0	0	0	0	0	0	4	0	0	0	0	0

Figura 7.4: Matriz de adyacencia de orden 15

15. Algebraic and Geometric Topology (ALGEBR GEOM TOPOL)

Como la dimensión es menor se muestra la matriz de adyacencia de orden 15 involucrada, ver Figura (9.4), se entiende que la posición está regida por la lista anterior.

El código se encuentra en el capítulo del Apéndice (A.6). Nuevamente mostramos el resultado final del código

Listing 7.2: Primer Algoritmo de Prueba

```
Iteraci n n mero:  9
Error relativo:  [-1.11022302e-16]
(array([10], dtype=int64),)
Error relativo alcanzado
se necesitaron 9 iteraciones
Vector de ordenamiento de menor a mayor:
[[ 7  9  5 13 14  8  6  2  1  4 12  3 15 10 11]]
```

7.4. Clasificación de autores

Continuamos con los pasos de la clasificación anterior y el modo de extracción de datos mencionada en el Apéndice. Comparando con las clasificaciones de artículos y revistas podemos mencionar:

1. Desde una visión general, esta clasificación puede considerarse una mixtura de las dos anteriores. Para Scopus los autores o investigadores pueden pertenecer a varias instituciones, pueden escribir en varias revistas; pero a su vez un autor puede escribir múltiples artículos a lo largo de su carrera profesional.

2. La comparación de la clasificación de artículos y autores, es que en general cuando comparamos artículos hay una mayor cantidad a clasificar comparado con la cantidad de autores; pero en contraparte los artículos tienen, o suelen tener, menos relaciones entre ellos, en cambio los autores van acumulando relaciones entre sus colegas a lo largo del tiempo.

3. La comparación de la clasificación de revistas y artículos, es que en general cuando comparamos revistas hay una menor cantidad a clasificar comparado con la cantidad de autores; pero en contraparte las revistas tienen, o suelen tener, más relaciones entre ellos, ya que la cantidad de autores que publican en una revista puede ser amplia.

Para esta clasificación usaremos la información de autores de la Universidad Mayor de San Marcos, en un período también de 5 años, desde el 2017, de Medicina de la subdisciplina de Oncología, encontramos un total de 36.

1. Autor 1
2. Autor 2
3. Autor 3
4. Autor 4
5. Autor 5
6. Autor 6
7. Autor 7
8. Autor 8
9. Autor 9
10. Autor 10
11. Autor 11
12. Autor 12
13. Autor 13
14. Autor 14
15. Autor 15
16. Autor 16
17. Autor 17
18. Autor 18
19. Autor 19
20. Autor 20
21. Autor 21
22. Autor 22
23. Autor 23
24. Autor 24

25. Autor 25
26. Autor 26
27. Autor 27
28. Autor 28
29. Autor 29
30. Autor 30
31. Autor 31
32. Autor 32
33. Autor 33
34. Autor 34
35. Autor 35
36. Autor 36

La dimensión de la matriz de adyacencia es 36, por lo que solo vamos a mostrar la información del filtro utilizado.

1	<	Top 500 authors, by Scholarly Output	
2	Entity	Universidad Nacional Mayor de San Marcos	
3	Year range	2017 to 2021	
4	Subject classification	ASJC	
5	Filtered by	Oncology	
6	Types of publications included	all publication types	
7			
8	Data source	Scopus	
9	Date last updated	7 September 2022	
10	Date exported	15 September 2022	
11			
12	Name	Scopus author profile	Scopus author ID
13		https://www.scopus.com/authi	57223224507
14		https://www.scopus.com/authi	56771393000
15		https://www.scopus.com/authi	35483049600
16		https://www.scopus.com/authi	57202975311
17		https://www.scopus.com/authi	24390696700
18		https://www.scopus.com/authi	57219380881
19		https://www.scopus.com/authi	57222260898
20		https://www.scopus.com/authi	35263574400

Figura 7.5: Filtro usado para la clasificación por importancia de autores.

El código se encuentra en el capítulo del Apéndice (A.7). Nuevamente mostramos el resultado final del código

Listing 7.3: Primer Algoritmo de Prueba

```
Iteraci n n mero:  23
Error relativo:  [-1.11022302e-16]
```

```
(array([31], dtype=int64),)
Error relativo alcanzado
se necesitaron 23 iteraciones
Vector de ordenamiento de menor a mayor:
[[ 1 24  6  2 17 16  3 19 14 26 33 22  9 28 11 10  7 29 30  4 27  8 15 13
  21 23 25 31 20  5 18 36 34 35 12 32]]
```

Capítulo 8

Resultados e interpretaciones

Resultado respecto a los artículos

Ordenación de importancia de artículos mostrado de mayor a menor.

1. genetic structure of drug-resistant mycobacterium tuberculosis strains in peru based on haplotypes obtained from a line probe assay
2. Artificial intelligence and innovation to optimize the tuberculosis diagnostic process
3. Molecular diagnosis of multidrug-resistant tuberculosis in sputum samples by melting curve análisis
4. Feasibility of an mobile application as a tool for multidrug-resistant tuberculosis contact monitoring in Peru
5. Cultural practices linked to health care and perception on the attention health facilities in residents of high-Andean settlements in Huancavelica, Peru
6. Clinical characteristics and plasma exchange response in guillain-barré patients
7. Clinical characteristics of pyrazinamide-associated hepatotoxicity in patients at a hospital in lima, peru
8. Infectious agents in biological samples from patients with guillain-barré syndrome in Peru, 2018-2019
9. Disease progression velocity as predictor of severity in guillain-barre síndrome
10. From ALMA-ATA to the digital citizen: Towards a digital primary health care in Peru

Interpretación respecto a los artículos

1. El que realiza el proceso de elección de la disciplina a clasificar (por ejemplo Medicina, o subdisciplina como el Álgebra) y la construcción de la matriz de adyacencia del tema, es decir, recabar las referencias recibidas (citas) y las referencias entregadas puede notar dos cosas:

a) La diferencia en la cantidad de artículos de ramas del conocimiento como Medicina, Ciencias Sociales, Ciencias humanas, etc., comparado con ciertas subdisciplinas de las Ciencias Naturales puede ser amplia. Esto se menciona en [14], pág. 13: "Los académicos que trabajan en diferentes disciplinas muestran características distintas en su enfoque de la investigación y en su comunicación sobre la investigación. Estas diferencias de comportamiento no son mejores o peores entre sí, sino que son simplemente un hecho asociado con campos particulares de investigación".

b) En cuanto a la relación entre artículos, esta es escasa. Esto puede deberse a muchos factores que pueden ir desde la naturaleza del campo de estudio, que puede ser para grupos reducidos, o respecto a nuestra coyuntura una falta de inversión en ciertos campos de investigación, etc. En [14], pág. 13, mencionan lo siguiente: "Los académicos en campos como la Ingeniería Química tienden a publicar con más frecuencia que los investigadores en Matemáticas. Las publicaciones en disciplinas como la Toxicología tienden a tener listas de referencias mucho más largas que las de las ciencias sociales. La investigación en física tiende a ser mucho más colaborativa que la investigación en artes y humanidades, lo que resulta en un mayor número de coautores por publicación".

2. Teniendo en cuenta el punto anterior, para ejemplificar el algoritmo se escogió una disciplina que en Perú goce con abundancia de artículos por clasificar, en este caso Medicina; pero de todas formas al recabar la información es notorio el comportamiento mencionado en 1.b

3. A pesar de que la relación entre los artículos no es abundante, los 4 primeros puestos muestran la relación que hay entre ellos.

   a) Se tratan sobre estudios relacionados a la tuberculosis.

   b) Sin embargo, el puesto 1 y 3 ya difieren de los puestos 2 y 4, ya que el primer par son estudios de índole médico y los otros apuntan a uso de herramientas tecnológicas como la Inteligencia Artificial o desarrollo de aplicaciones móviles para el monitoreo, respectivamente.

4. Luego, se ve también una relación entre los puestos 6, 8 y 9, son artículos relacionados al Síndrome de Guillain-Barré, aunque esta relación es más íntima comparado con las que guardaban los 4 primeros puestos.

5. Para el resto de artículos de estos 10 primeros puestos, no hay una relación clara entre ellos, caso de no haberla. Una correcta interpretación de tales posiciones requiere un análisis interpretativo de los subsiguientes puestos, pero también amerita la pericia de una opinión formada en tal materia, la medicina.

6. La relación entre ciertos artículos de los 10 primeros puestos de este ordenamiento son un punto a favor del desarrollo del modelo.

Resultado respecto a las revistas

Ordenación de importancia de artículos mostrado de mayor a menor.

1. Annals of Mathematics (ANN MATH)
2. Compositio Mathematica (COMPOS MATH)
3. Algebraic and Geometric Topology (ALGEBR GEOM TOPOL)
4. Journal of pure and Applied Algebra (J PURE APPL ALGEBRA)
5. Algebra and Number Theory (ALGEBR NUMBER THEORY)
6. Algebras and Representation Theory (ALGEBRA REPRESENT TH)
7. Journal of Algebra (J ALGEBRA)
8. Advances in Mathematics (ADV MATH)
9. Journal of Group Theory (J GROUP THEORY)
10. Algebra Colloquium (ALGEBR COLLOQ)
11. Journal of Commutative Algebra (J COMMUT ALGEBR)
12. Journal of Algebra and its Applications (J ALGEBRA APPL)
13. Communications in Algebra (COMMUN ALGEBRA)
14. Linear Algebra and its Applications (LINEAR ALGEBRA APPL)
15. Linear and Multilinear Algebra (LINEAR MULTILINEAR A)

Interpretación respecto a los artículos

1. Para este ordenamiento se hizo un filtro previo, de 310 revistas indexadas sobre matemática se tomaron 15 relacionadas al Álgebra; a pesar de tener pocas revistas las relaciones entre los artículos que conforman una revista se ven reflejadas en la matriz asociada.
2. Los puestos 3, 5 y 6 son de temática similar, una aplicación común del Álgebra Conmutativa y en otros ránquines muestran puestos similares. Sin embargo, hay puestos disímiles respecto a otros ránquines, por ejemplo, el puesto 10 y 11 de este ranking lidera otros, modelados de otra manera, por ejemplo, el conocido Factor de Impacto (FI).
3. Esta variabilidad de puestos en diferentes ránquines no es más que una manifestación de la manera que se modelan estos; antiguos métricas no toman en consideración la correlación que hay entre ellas, en cambio nuestro modelo sí.

Resultado respecto a los autores

Ordenación de importancia de autores mostrado de mayor a menor.

1. Autor 32
2. Autor 12

3. Autor 35
4. Autor 34
5. Autor 36
6. Autor 18
7. Autor 5
8. Autor 20
9. Autor 31
10. Autor 25
11. Autor 23
12. Autor 21
13. Autor 13
14. Autor 15
15. Autor 8
16. Autor 27
17. Autor 4
18. Autor 30
19. Autor 29
20. Autor 7
21. Autor 10
22. Autor 11
23. Autor 28
24. Autor 9
25. Autor 22
26. Autor 33
27. Autor 26
28. Autor 14
29. Autor 19
30. Autor 3
31. Autor 16
32. Autor 17

33. Autor 2
34. Autor 6
35. Autor 24
36. Autor 1

Interpretación respecto a los autores

1. A comparación de los casos de revistas y artículos, estamos clasificando autores, por lo que una correcta interpretación de los resultados tiene que venir de la mano de un conocimiento del trabajo de los autores.
2. Sin embargo, la relación que existe entre los autores y es vista en la matriz de adyacencia en la extracción de datos es mayor comparada con el ejemplo de los artículos, y aunque menor comparado con el de revistas, donde por ejemplo una revista referencia a otra hasta 174 veces, aquí hay autores que se referencian entre ellos 21 veces; esto en el lapso de los últimos 5 años.

Capítulo 9

Conclusiones y recomendaciones

9.1. Conclusiones

Conclusiones generales

1. Se ha modelado y ejecutado un algoritmo que como resultado da una ordenación por importancia de los objetos involucrados. Estos objetos pueden ser artículos, revistas, autores.

2. Para tal ordenación no es necesario que tales objetos sean obtenidos con alguna métrica previa, lo que se necesita es tener información de como estas se relacionan entra ellas. Ciertamente, se está creando una métrica en el nicho que se analice, pero estos son de índole ordinal: primero en importancia, segundo en importancia, etc.

3. La posición de un objeto en la ordenación es debido a dos factores: La cantidad de relaciones entre objetos (la referenciación) y la calidad de los mismos (no es lo mismo se referenciado por algunos objetos que a su vez son pocos referenciados, que ser referenciado por un objeto que a su vez es muy referenciado); esta interconexión es la esencia del algoritmo.

4. El modelamiento planteado está basado en el algoritmo PageRank, y muchas fuentes la recomiendan como un competidor directo de muchas métricas bibliográficas.

5. A comparación de otras métricas, al considerar la importancia de un objeto directamente proporcional a la suma de las importancias de los objetos que la referencian, nuestro modelo gana en robustez y ciertamente los principales factores que pueden viciar sus resultados son:

   a) La constante α en la etapa de modelamiento, algunos autores escogen valores similares al utilizado acá (0.85), y esto puede repercutir en puestos, sobre todo intermedios.

   b) Escoger objetos para el ordenamiento que no tengan disciplinas similares.

   c) Información faltante que contenga pocos datos (referenciaciones) puede tener un impacto negativo en el posicionamiento de ciertos objetos.

6. En contraparte a lo mencionado previamente, el modelo muestra ventajas respecto aspectos como la poca conexidad que puedan tener los objetos, por ejemplo, el caso de los artículos en este trabajo. Ya que a pesar que la base de datos en la que se

trabaje proporcione datos con poca conexidad, el ranking muestra que ciertos puestos logran mostrar una relación.

7. Comparado con otras métricas, nuestro modelo no considera las autorreferencias, lo que muestra solidez en sus resultados. La preocupación de las autocitas excesivas que inflan artificialmente el posicionamiento de, por ejemplo, revistas, son consideraciones que tienen que prever muchos modelos.

Conclusiones específicas

1. De los artículos, y autores. El modelo asegura solidez en sus resultados, a pesar de quizás tener poca conexidad en los datos involucrados, sin embargo, es conveniente tener en cuenta los lapsos de tiempos involucrados; diversas métricas aconsejan un lapso no menor de 4 y no mayor de 7.

2. De los artículos, y autores. Algunas métricas adoptan medidas para ganar robustez en el modelo, por ejemplo, considerar artículos o autores que hayan sido referenciados más de 12 veces, incluso en lapsos de tiempos reducidos. Lo puede ser una buena praxis, depende mucho del nicho que se busque ordenar.

3. Al buscar disciplinas en las que poder utilizar nuestro modelo, el principal inconveniente fue buscar esta robustez, por ejemplo, depende mucho de la cantidad de investigadores que pueda tener un país o institución.

4. De las revistas. El modelo PageRank es quizás el competidor por excelencia de diversas métricas en la actualidad, como es por ejemplo el Factor de Impacto (FI) del JCR (Journal Citation Reports), ya que la conexidad entre revistas suele ser alta, y la investigación científica se apoya fuertemente en este tipo de documentos por lo que la información recabada del campo o disciplina puede corresponderse en gran medida con la realidad, lo que se requiere es un modelo que interpreta correctamente esa información.

5. De las revistas. Al comparar nuestro modelo y el FI, podemos ver dos diferencias esenciales:

   a) El FI tiene en consideración un intervalo de tiempo menor de la data, 2 años, en cambio nosotros en general planteamos un promedio de 4 o 5 años.

   b) El FI tiene en su modelo aspectos distintos al nuestro, sobre todo en cierta parte del proceso requiere esta formulación: $a_i j = c_{ij}/(p_{1i} + p_{2i})$, en cambio el nuestro considera la importancia de un objeto directamente proporcional a la suma de las importancias de los objetos que la referencian.

   c) Esta última idea encapsula en nuestro modelo al Teorema de Perron - Frobenius. Herramienta que se volvió fundamental en diversos campos.

9.2. Recomendaciones

1. El trabajo tiene entre sus motivaciones que el lector pueda entender la *lógica* que subyace el modelamiento del algoritmo, en ese sentido es recomendable una vez entendido el funcionamiento el lector pueda modificar ciertos coeficientes, sobre todo α, de acuerdo a la necesidad de su uso.

2. El lector interesado en obtener información más específica y de manera sistemática de SCOPUS puede acceder a las APIS de la web, aunque muchas de sus funcionalidades están restringidas para usuarios con suscripción, aún los usuarios de menor jerarquia tienen funcionalidades interesantes.

3. El código fue escrito en **Python**, una manera rápida de adaptarse al modelo puede consistir en rescribir o adaptar al código a otros lenguajes, esto sobre todo para instituciones que promueven, o debido a necesidad de presupuestos, utilizan softwares libres.

4. El método de las potencias no es el único método que permite calcular la ordenación, existen otros métodos convenientes de acuerdo a la potencia de cómputo del usuario.

5. Un resultado indirecto, pero importante del trabajo es incentivar un correcto almacenamiento de los datos que la institución del lector vaya produciendo, ya que esta data más temprano que tarde siempre debe ser analizada.

6. La necesidad de contar con métricas (u ordenaciones) viene, generalmente, del hecho de tomar de decisiones. Para tal fin, es siempre recomendable contar con más de una métrica y tener una idea de cómo trabajan.

7. Las métricas más adecuadas siempre dependerán de la pregunta particular que el usuario esté haciendo. El mejor enfoque es resaltar algunos puntos clave que es importante tener en cuenta y que el usuario aplique su sentido común.

Apéndice A

Apéndice

A.1. Desarrollo de los fundamentos matemáticos

A.1.1. Teoría de Grafos

La teoría de grafos empezó en 1736 cuando Leonhard Euler resolvió el muy conocido problema de los siete puentes de Königsberg. Este problema requería obtener un camino circular por la ciudad de Königsberg (ahora Kaliningrado) de tal manera que cruzara cada uno de los siete puentes una sola vez, para tal propósito Euler notó que la teoría que resuelva el ejercicio dependía solo de sus propiedades de conexión. Los territorios envueltos se pueden representar por puntos (llamados vértices), y los puentes por curvas que unen los respectivos puntos; de tales representaciones obtenemos un grafo.

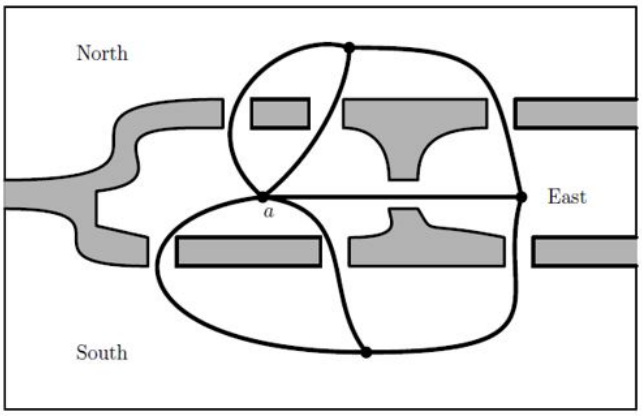

Figura A.1: Problema de los puentes de Konigsberg

La teoría de grafos en este trabajo es la teoría inicial para relacionar objetos como son artículos, autores, revistas; para tal fin un grafo es una estructura matemática que representa problemas cotidianos de manera gráfica. A lo largo del texto vamos poner énfasis en el siguiente ejemplo: Si tuviéramos que valorar 4 artículos, los cuales algunos hacen referencia al trabajo de otro(s), para tener un panorama de este problema el primer paso natural es plasmar en una hoja de papel a los 4 artículos con alguna letra y un punto, quizás a,b,c,d, luego para entender gráficamente que un artículo hace referencia a otro podemos dibujar una flecha dirigida, así obtendríamos una grafo muy parecido al de la Figura 1.1. Por tal motivo esta teoría aparece de manera natural al tratar este problema.

Definición A.1.1. *Un grafo es un par* $G = (V, E)$ *consistente de un conjunto vinito* $V \neq \emptyset$

y un conjunto E de subconjuntos de dos elementos (pares) de V. Los elementos de V son llamados vértices. Un elemento $e = \{a, b\}$ de E es llamado una arista con vértices finales a y b.

Para muchas aplicaciones - especialmente problemas concernientes a tráfico y transporte - es muy útil darles una dirección a las aristas de un grafo.

Definición A.1.2. *Un grafo dirigido es un par $G = (V, E)$ consistente de un conjunto vinito $V \neq \emptyset$ y un conjunto E de pares ordenados de V. Los elementos de V son llamados vértices. Un elemento $e = (a, b)$ de E es llamado arista dirigida que une los vértices a, b y se denota $a \to b$.*

Si tenemos los vértices (a, b) y (b, c), notar que bajo la idea de grafo dirigido la noción de ir desde a hacia c es clara, iniciando con (a, b) y luego (b, c), esto puede ser llamado un camino desde a hacia c.

A continuación, se define una noción de conectividad en los grafos. Esta propiedad es importante en diversos campos de aplicación de la teoría de grafos, permite describir que tan conectados están los vértices del grafo ya que de tener muchos vértices existe la posibilidad de tener ciertos vértices que se conecten solamente entre ellos, originando que el grafo pueda verse como varios grafos totalmente aislados, esto tiene sus pros y contras. Ese método se conoce como particionar el grafo, en nuestro caso se desea que nuestro grafo tenga esta conexidad, es decir que nuestro grafo no pueda dividirse.

Definición A.1.3. *Un grafo dirigido G es llamado fuertemente conexo, si para todo par de vértices a, b hay algún camino de a hacia b y otro de b hacia a.*

Ejemplo 1. Es sencillo comprobar que el grafo G con $V = \{a, b, c\}$ y $E = \{(a, b), (b, c)\}$ no es fuertemente conexo.

Ejemplo 2. Es sencillo comprobar que el grafo G con $V = \{a, b, c\}$ y $E = \{(a, b), (b, a), (b, c), (c, b)\}$ es fuertemente conexo.

Para este estudio los vértices pueden representar autores, artículos, etc., y las aristas dirigidas la relación que hay entre ellos. Cuando se poseen pocos datos uno puede plasmar con papel y lápiz un grafo, si el volumen de datos aumenta puede usar softwares que permitan visualizar tales grafos, pero cuando el volumen es desmedido se sugiere plasmar la información del grafo en una matriz.

A.1.2. Teoría de matrices

Una matriz es una tabla compuesta por números, llamados entradas, en general reales o complejos, los cuales se ordenan en m filas y n columnas (dimensión $m \times n$). Se denota:

$$A = (a_{ij}), \quad \text{donde } i = 1, \ldots, m; j = 1, \ldots, n; a_{ij} \in \mathbb{R}(o\ \mathbb{C})$$

En el conjunto de matrices se permite definir operaciones naturales entre matrices, suma de matrices, producto de matrices y producto de un número por una matriz, siempre y cuando la dimensión sea *compatible*. Para nuestros propósitos es claro que la primera y tercera operación mencionadas hacen del conjunto de matrices de orden $m \times n$ un *espacio*

vectorial.

Esta teoría permite tener un orden en cuanto a la cantidad de objetos a tratar (artículos, autores o revistas) ya que al ponerlos en una tabla podemos considerar la relación entre ellos de una manera más ordenada. También tiene importancia por su utilidad en representar funciones entre espacios vectoriales y representar sistemas de ecuaciones lineales.

Empezamos con la equivalencia que hay entre un grafo dirigido con n vértices y una matriz de orden n.

Definición A.1.4. *Sea $G = (V, E)$ un grafo dirigido con n vértices, la matriz de adyacencia $A(G) = (a_{ij})$ de G es la matriz definida por*

$$a_{ij} = \begin{cases} 1, \text{ Si existe } i \to j \text{ en } G \\ 0, \text{ para otros casos} \end{cases}$$

A partir de ahora la teoría de matrices se muestra extremadamente útil. Por ejemplo, un grafo de 4 vértices se puede representar en una matriz cuadrada de orden 4. La relación entre los vértices se refleja en el valor de las entradas de la matriz. la utilización de esta matriz en el modelo se encuentra en el capítulo 4.

A continuación, algunos tipos de matrices:

Definición A.1.5. *Una matriz $A = (a_{ij})$ de orden $m \times n$ se dice no negativa si todas sus entradas son mayores o iguales que 0, $a_{ij} \geq 0$, y se denota $A \geq 0$. Si todas las entradas son positivas se denota $A > 0$, y se le llama matriz positiva.*

Cuando una matriz tiene una columna ($n = 1$) se denomina vector de dimensión m. Se hace esto porque en la teoría de espacios vectoriales V, las funciones que respetan la estructura vectorial y que son llamados aplicación lineal, función lineal, transformación lineal, u operador lineal pueden ser representados por matrices.

Ejemplo 3. Sea $T : R^2 \to R^3$ una transformación lineal definida por $T(x, y) = (3x + y, 2y, -x + 5y)$. Se tiene dos espacios vectoriales, uno bidimensional y otro tridimensional, además de una función T. El espacio bidimensional puede ser visto como un conjunto de matrices de orden 2×1, de manera análoga el otro espacio como un conjunto de matrices de orden 3×1. Así, la función T se puede representar como una matriz de orden 3×2, de la siguiente manera:

$$\begin{bmatrix} 3 & 1 \\ 0 & 2 \\ -1 & 5 \end{bmatrix}$$

Ya que si multiplicamos T por $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$, obtendremos la transformación mencionada previamente, dicho producto puede denotarse como $T \cdot v$ o $T(v)$, según sea conveniente.

A partir de ahora un espacio vectorial de dimensión m será considerado como el espacio de matrices con m filas y 1 columna con entradas reales, es decir $V = \mathbb{R}^m$. Una transformación lineal que va desde $\mathbb{R}^n$ a $\mathbb{R}^m$ es un matriz de orden $m \times n$. Por lo tanto, una matriz puede representar una transformación lineal, y una matriz columna puede representar un vector.

Los párrafos siguientes tienen su importancia en la demostración del teorema más importante del trabajo, el Teorema de Perron-Frobenius.

Ahora una propiedad de ciertos conjuntos que se mencionará más adelante es la de invarianza respecto a una transformación lineal T.

Definición A.1.6. *Un subconjunto $S \subseteq V$ es invariante respecto a T si $T(S) \subseteq S$.*

En general se considera S como un subespacio vectorial de V. Un tipo de vector asociado a la invarianza es el siguiente.

Definición A.1.7. *Dada una transformación línea $T : V \to V$, un número $\lambda \in \mathbb{R}$ se llama autovalor (o valor propio) de T, si existe un vector $v \in V$, $v \neq 0$, tal que $T(v) = \lambda v$. Este vector v se llama autovector (o vector propio) de T correspondiente al autovalor λ.*

Para calcular los autovalores se resuelve la ecuación $P(\lambda) = \texttt{det}(A - \lambda I) = 0$, luego para los respectivos autovectores la ecuación $(A - \lambda I)v = 0$. Cuando se calculan los autovalores, que son raíces del polinomio $P(\lambda) = 0$, estos vienen acompañados de una multiplicidad, que será denominada *multiplicidad algebraica*.

Definición A.1.8. *Dado $\lambda \in \mathbb{R}$ valor propio de la matriz A, llamamos subespacio propio asociado a λ al conjunto $E_\lambda = \{x \in \mathbb{R}^n : \ A \cdot x = \lambda \cdot x\}$, es decir a todos los vectores propios asociados a λ, junto con el vector nulo 0.*

Definición A.1.9. *Dado $\lambda \in \mathbb{R}$ valor propio de la matriz A, llamamos multiplicidad geométrica de λ a la dimensión del subespacioE_λ, es decir, al número mínimo de parámetros necesarios para expresar la solución general del sistema $(A - \lambda I) \cdot x = 0$.*

Propiedad 1. Los autovectores de una transformación lineal $T : V \to V$ satisfacen:

1. Vectores propios asociados a valores propios distintos son linealmente independientes.
2. Todo vector propio está asociado a un único valor propio.

Definición A.1.10. Sean A y B dos matrices cuadradas del mismo orden. Se dice que A y B son semejantes, si existe una matriz C invertible tal que $A = C \cdot B \cdot C^{-1}$.

En general, toda matriz es semejante a uno de forma triangular (superior, inferior, diagonal). De este tipo de matrices la más simple recibe el nombre de *forma canónica de Jordan*. Para su entendimiento tengamos en cuenta lo siguiente.

Definición A.1.11. *Una matriz A es nilpotente si existe un número entero $q > 1$ tal que $A^q = 0$. El menor número q es llamado índice de nilpotencia.*

Es decir, hay una potencia de A que es la matriz nula, por ende, la transformación lineal T asociada a la matriz A al ser compuesta (aplicada) consigo mismo q-veces se convertirá en la función nula.

Así como la propiedad de nilpotencia está presente en la matriz y en la transformación lineal, lo mismo ocurre para la inversibilidad, si la matriz es inversible, también lo es la transformación lineal.

A continuación dos propiedades que fundamentan la forma canónica de Jordan.

Propiedad 2. Dada una transformación líneal $T : V \to V$, existen subespacios U y W invariantes por T, tales que

1. $V = U + W$, además $U \cap W = 0$.

2. T restringido a U es nilpotente y restringido a W es inversible.

Esta descomposición de T en sus partes nilpotente e inversible es *única*.

Propiedad 3. Forma Canónica de Jordan. Sea V un $\mathbb{C}$-espacio vectorial de dimensión finita y $T : V \to V$ una transformación lineal, $\lambda_1, \ldots, \lambda_r$ sus distintos autovalores y $m_1, \ldots, m_r$ sus correspondientes multiplicidades algebraicas, ordenadas en la forma

$$m_1 \geq m_2 \geq \ldots \geq m_r \geq 1$$

Entonces existen subespacios V_j, $j = 1, \ldots, r$ invariantes por T, tales que:

1. $V = V_1 + \cdots + V_r$, y además $V_i \cap V_j = 0$, cuando $i \neq j$.

2. $\dim V_j = m_j$, $j = 1, \ldots, r$.

3. $T - \lambda_j I$ es nilpotente sobre V_j.

A continuación, dada una matriz A hallaremos la matriz triangular semejante a ella.

Ejemplo 4. Dada la matriz

$$A = \begin{bmatrix} 7 & 1 & -9 \\ -1 & 0 & 0 \\ 5 & 1 & 7 \end{bmatrix},$$

su polinomio característico viene dado por $f_A(x) = \det(xI - A) = (x+1)^2(x-2)$, que tiene raíces (autovalores) $\lambda = -1$ de multiplicidad dos, y $\alpha = 2$.

Para el autovalor $\lambda = -1$ de multiplicidad dos se plantea la ecuación $Av = (-1)v$, se obtiene la solución $v = (1, 1, 1)^t$, que genera un subespacio vectorial invariante de dimensión 1, que no coincida la multiplicidad algebraica y la dimensión del subespacio significa que la matriz no es semejante a una matriz diagonal. Hay otro vector correspondiente a este autovalor, planteando $(A - (-1)I)w = v$, se obtiene la solución $w = (0, 1, 0)^t$.

Para el autovalor $\alpha = 2$, se plantea $Au = 2u$, y se obtiene la solución $u = (2, -1, 1)^t$.

Así se tiene la base $\{v, w, u\}$ de $\mathbb{R}^3$, cuyas componentes determinan la matriz P de cambio de base

$$P = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix},$$

observe que está formado por vectores columna, que son soluciones de los sistemas anteriores, además cuya inversa es

$$P^{-1} = \begin{bmatrix} -1 & 0 & 2 \\ 2 & 1 & 3 \\ 1 & 0 & -1 \end{bmatrix}$$

Al hacer el siguiente producto, podemos ver la forma canónica de Jordan:

$$P^{-1}AP = J(A) = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

En esta matriz vemos dos bloques: $\left[\begin{smallmatrix} -1 & 1 \\ 0 & -1 \end{smallmatrix}\right], [2]$; en cuya diagonal están los autovalores correspondientes. Observe que el primer bloque induce una transformación lineal entre espacios bidimensionals, el segundo bloque una multiplicación por 2, y aunque suene trivial, una transformación lineal entre espacios unidimensionales.

Los bloques que interesan en una demostración en este trabajo son los bloques de dimensión dos del tipo:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}, \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

donde $\lambda \neq 0, \beta \neq 0, \lambda$ arbitrario.

Definición A.1.12. *La matriz permutación es la matriz cuadrada con todos sus $n \times n$ elementos iguales a 0, excepto uno cualquiera por cada fila y columna, el cual debe ser igual a 1.*

Ejemplo 5. Para matrices de orden 3, algunos casos de matriz permutación son:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Ahora la matriz más importante que se utilizará, la matriz irreducible, y está relacionada con el grafo fuertemente conexo:

Definición A.1.13. *Una matriz A de orden $n \times n$, con $n \geq 2$, se dice que es reducible si existe una matriz permutación P tal que:*

$$P^t AP = \begin{bmatrix} A_1 1 & A_1 2 \\ 0 & A_2 2 \end{bmatrix}$$

donde $A_1 1$, $A_2 2$ son matrices cuadradas de orden menor que n. ***Si no existe tal matriz de permutación P, entonces diremos que A es irreducible.***

Ejemplo 6. La falta de entradas nulas en la matriz se debe entender en el hecho que el grafo asociado tiene muchas aristas dirigidas por ende tenemos ?mucha conexidad?. Veamos algunos casos.

- Toda matriz positiva es irreducible.
- Toda matriz no negativa 3×3 que tiene sólo una entrada nula, es irreducible.
- Toda matriz no negativa $n \times n$ con $n \geq 2$ que tiene a lo más $n - 2$ entradas nulas, es irreducible.

Finalmente, un tipo de matriz importante para el trabajo, su ventaja es de índole computacional, a medida que sus potencias crecen, estas poseen un vector conocido como estado estacionario que se comporta como un autovector de autovalor 1.

Definición A.1.14. *Una matriz A cuadrada de orden n es estocástica si sus entradas $a_{ij} \in \mathbb{R}$ cumplen lo siguiente:*

1. $a_{ij} \geq 0$.
2. *La suma de las entradas de cada columna es 1.*

A.1.3. Teoría de Topología

La topología está dedicada al estudio de las propiedades de los cuerpos geométricos que permanecen inalteradas por funciones continuas. Una de las tantas ideas intuitivas que tiene la topología es ¿Cómo es el *suelo* matemático en el que estamos trabajando? Para este estudio se definen ideas puntuales para poder acceder al famoso Teorema del punto fijo de Brouwer.

La importancia de esta teoría en este trabajo es de índole técnico, y sirve para poder utilizar un teorema en la demostración del Teorema de Perron-Frobenius.

Definición A.1.15. *Sea $X \neq \emptyset$ y $P(X)$ el conjunto potencia de X (conjunto formado por todos los subconjuntos de X). Se dice que $\tau \subset P(X)$ es una topología sobre X sí y solo si:*

1. $\emptyset, X \in \tau$
2. *Si* $U_1, U_2 \in \tau \Rightarrow U_1 \cap U_2 \in \tau$
3. *Si* $\{U_\lambda\} \subset \tau \Rightarrow \bigcup_\lambda U_\lambda \in \tau$

Si τ es una topología sobre X entonces los elementos de τ son llamados conjuntos τ-abiertos o simplemente abiertos. Se dice que el par (X, τ) es un espacio topológico si y sólo si $X \neq \emptyset$ y $\tau \subset P(X)$ es una topología sobre X.

Ejemplo 7. El conjunto $\mathbb{R}^2$ tiene una topología natural conformada por todos los discos abiertos, de cualquier radio y cualquier centro. Verificar esto no es un ejercicio sencillo, la manera más fácil de entender la topología en $\mathbb{R}^2$ es comprendiendo el concepto de *espacio métrico.*

Se sigue con las definiciones.

Definición A.1.16. *Se dice que un subconjunto de A en un espacio topológico X es un retracto de X si existe una aplicación continua $r : X \to A$ (llamada retracción) tal que $r(a) = a$ para todo $a \in A$.*

Definición A.1.17. *La bola unitaria cerrada B^n en $\mathbb{R}^n$ es el conjunto $B^n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$, donde $\|x\| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$.*

Definición A.1.18. *Sea f una función de un conjunto A en sí mismo. Se dice que $x \in A$ es punto fijo de f si $f(x) = x$.*

No todas las funciones tienen puntos fijos. Por ejemplo, si f es una función definida sobre los números reales como $f(x) = x + 1$, entonces f no tiene ningún punto fijo, ya que x no es nunca igual a $x + 1$ para ningún número real.

Nótese que esta definición es cercana a la invarianza, solo que aquí se trata de un elemento, un conjunto unitario.

Ejemplo 8. Las siguientes funciones tienen puntos fijos. Para obtenerlos hay que resolver la ecuación $f(x) = x$.

1. $f : \mathbb{R} \to \mathbb{R}, f(x) = -x$, cuenta con un solo punto fijo.
2. $f : \mathbb{R} \to \mathbb{R}, f(x) = x^3$, cuenta con 3 puntos fijos.
3. $f : [0, 10] \to \mathbb{R}, f(x) = x^3$, cuenta con 2 puntos fijos.
4. $f : \mathbb{R} \to \mathbb{R}, f(x) = x$, cuenta con infinitos puntos fijos.

El teorema a continuación es pieza clave en la demostración del Teorema de Perron-Frobenius.

Veamos la idea que hace utilizar este teorema: Si tenemos 4 artículos que se desea ordenar por importancia, ya sabemos que para visualizar las relaciones entre ellos necesitamos una matriz de adyacencia, más adelante se verá que también se necesita un vector columna. Así, una entrada del vector columna representa un artículo, para poder ordenarlos este se hará a través del valor que tenga cada entrada, por ejemplo, si obtenemos un vector con las entradas $(0{,}4, 0{,}5, 0{,}3, 0{,}7)$, se entiende que el que tiene mayor importancia es el cuarto artículo con 0.7, el segundo en importancia es el segundo artículo con 0.5, el tercero en importancia es el primer artículo con 0.4 y el último en importancia es el tercer artículo con 0.3.

Como ya se dijo, los vectores del espacio $\mathbb{R}^n$ se pueden considerar como matrices columna, si deseamos obtener coordenadas positivas es obligatorio movernos en la parte del espacio que solo tenga coordenadas positivas, llamémosle R^+. Cierta transformación lineal A, vista como matriz, se aplicará en vectores del conjunto R^+, en la demostración se asegura que el conjunto R^+ es invariante por A, pero más importante aún el Teorema del Punto fijo asegura *la existencia de un vector invariante*, **que, bajo el punto de vista de nuestro trabajo, tendrá en sus coordenadas valores positivos que sirve para la ordenación por importancia**.

Se enuncia el teorema de interés bajo dos enfoques, pero en esencia lo mismo.

Teorema A.1.19. ***Teorema del Punto fijo de Brouwer.*** *Si $f : B^n \to B^n$ es una función continua, entonces existe un punto $x \in B^n$ tal que $f(x) = x$.*

La convexidad, compacidad de un conjunto es una propiedad muy conocida en topología, los conjuntos que se usan en este trabajo cumplen las condiciones para tener estas propiedades y son sencillas de comprobar por lo que no ahondamos en eso, simplemente se da por sentado que estos conjuntos no son problemáticos y aceptan las hipótesis de nuestros teoremas.

Teorema A.1.20. ***Teorema del Punto fijo de Brouwer****. En un espacio euclídeo, toda aplicación continua de un conjunto convexo, compacto y no vacío, en sí mismo, admite un punto fijo.*

El conjunto R^+, mencionado previamente, tiene la propiedad de ser convexo, compacto y no vacío, por lo que cumple con las hipótesis del teorema y será usado en la demostración del Teorema de Perron-Frobenius.

A.1.4. Método de las potencias

El algoritmo está basado en la existencia de un autovector asociado al mayor autovalor positivo (módulo máximo), este vector es el que nos muestra el orden de importancia. Sin embargo, la demostración, en toda literatura, es no constructiva, es decir, establece como es que un objeto matemático con una cierta propiedad existe sin explicar como tal objeto se puede encontrar. Esta primera dificultad es resulta por este método.

El método de las potencias es un método iterativo utilizado para la obtención del autovalor de modulo máximo y su correspondiente autovector que nos asegura el Teorema de Perron-Frobenius.

La idea es atribuir una aproximación inicial arbitraria para el autovector correspondiente al autovalor dominante (de mayor módulo o tamaño) que es sucesivamente mejorado hasta que la precisión requerida sea encontrada. La convergencia para el autovalor dominante es simultáneamente obtenida.

Propiedad 4. Sea A una matriz cuadrada de orden n y la cual tiene n autovectores linealmente independientes y un autovalor estrictamente dominante, es decir, si sus valores propios son:

$$\lambda_1, \lambda_2, \ldots, \lambda_n$$

entonces ocurre que:

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$$

No hay pérdida de generalidad en suponerlo ordenados, de forma que el mayor módulo se corresponde con λ_1.

Si v_i es al autovector asociado a cada valor propio λ_i para $i = 1, 2, \ldots, n$ entonces por hipótesis tenemos que: $\{v_1, v_2, \ldots, v_n\}$ son linealmente independientes. Es decir, cualquier vector v de $\mathbb{R}^n$ se puede expresar como combinación lineal de los vectores $v_1, v_2, \ldots, v_n$, por lo que se tendrán $a_i \in \mathbb{R}$ tal que:

$$v = \sum_{n=1}^{n} a_i v_i$$

pero como v_i son autovectores y multiplicando (aplicando) la matriz A:

$$Av = A\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i A v_i = \sum_{i=1}^{n} a_i \lambda_i v_i$$

a esta sumatoria de vectores se le puede volver a aplicar la matriz A:

$$A^2v = A\left(\sum_{i=1}^{n} a_i\lambda_i v_i\right) = \sum_{i=1}^{n} a_i\lambda_i Av_i = \sum_{i=1}^{n} a_i\lambda_i\lambda_i v_i = \sum_{i=1}^{n} a_i\lambda_i^2 v_i$$

tenemos la siguiente expresión:

$$A^2v = a_1\lambda_1^2 v_1 + a_2\lambda_2^2 v_2 + \cdots a_n\lambda_n^2 v_n.$$

Si de manera sucesiva seguimos multiplicando la última expresión, se tendrá:

$$A^kv = a_1\lambda_1^k v_1 + a_2\lambda_2^k v_2 + \ldots a_n\lambda_n^k v_n.$$

Ya que tenemos un autovalor de mayor módulo (diferente de cero), podemos factorizarlo:

$$A^kv = \lambda_1^k\left(a_1v_1 + a_2\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \ldots a_n\left(\frac{\lambda_n}{\lambda_1}\right)^k v_n\right), \tag{A.1}$$

este proceso puede seguir *Ad infinitum*, ya que si se observa el proceso iterativo recae sobre los coeficientes, una división que involucra un denominador mayor que los numeradores, es decir se tiene el siguiente límite que no es otra cosa que multiplicar por sí mismo muchas veces un número menor que uno:

$$\lim_{n\to\infty}\left(\frac{\lambda_2}{\lambda_1}\right)^k = 0,$$

en consecuencia, el límite puede ser aplicado a la combinación lineal previa:

$$\lim_{k\to\infty} A^kv = \lim_{k\to\infty}\lambda_1^k\left(a_1v_1 + a_2\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \ldots a_n\left(\frac{\lambda_n}{\lambda_1}\right)^k v_n\right)$$

$$= \lim_{k\to\infty}\lambda_1^k \lim_{k\to\infty}\left(a_1v_1 + a_2\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \ldots a_n\left(\frac{\lambda_n}{\lambda_1}\right)^k v_n\right)$$

donde el segundo factor tiende a 0 a medida que k crece, es decir:

$$A^kv \approx \lambda_1^k a_1v_1.$$

En términos de sucesiones se puede expresar como:

$$\left\{\lambda_1^{-k}A^kv\right\} \longrightarrow \{a_1v_1\}.$$

Así a medida que vamos multiplicando por la matriz A vamos consiguiendo de manera más exacta el vector buscado, con una velocidad que depende del cociente entre los dos primeros autovalores.

A.2. Demostración matemática

El teorema más importante y que permite el desarrollo del algoritmo es el siguiente:

Teorema A.2.1. ***Teorema de Perron-Frobenius****. Sea A una matriz cuadrada de orden n, no negativa e irreducible, entonces existe un autovalor positivo y simple λ de A que tiene asociado un autovector positivo (es decir, con todas sus coordenadas positivas), además el valor absoluto de este autovalor es el mayor entre los valores absolutos de todos los autovalores de A.*

Demostración. Consideramos R^+ como el conjunto de rayos no negativos, es decir, las semirrectas con origen en el origen de coordenadas que están contenidas en el cuadrante C^+, definido como:

$$C^+ = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \ ; \ x_i \geq 0, \ i = 1, \ldots, n\}$$

(i) A es una transformación lineal invariante en el conjunto R^+ de rayos no-negativos, $A(R^+) \subset R^+$. Es decir, si A se evalúa sobre un rayo no negativo devuelve otro rayo no negativo. Esto ocurre porque tanto los escalares de la matriz como las coordenadas del rayo son elementos no negativos, entonces, cada coordenada del rayo se puede multiplicar por un escalar o sumarle un múltiplo de otra coordenada, pero nunca se va a volver negativa.

Es más, ningún rayo en R^+ se transforma en el elemento nulo.

Lo probamos por reducción al absurdo[1].

Supongamos que al evaluar $v = (0, \ldots, x_i, \ldots, 0)$ mediante A obtenemos un vector nulo, es decir, $v' = Av^t = (a_{1i}x_i, a_{2i}x_i, \ldots, a_{ni}x_i) = (0, \ldots, 0)^t$ obliga a que la columna i-ésima de A sea nula, entonces A es reducible (ya que esto implica que el grafo asociado a la matriz A no tiene ninguna arista llegando a la i-ésima arista, entonces desde ninguna arista se puede llegar a la i-ésima arista, por tanto el grafo asociado no es fuertemente conexo, o lo que es lo mismo, la matriz sería reducible).

Si en vez de tener sólo una coordenada no nula tuviese m, como tanto los elementos del vector como los de la matriz son no negativos, esto conllevaría que cada coordenada del vector imagen es el resultado de m sumandos no negativos, obliga a que los m sumandos sean nulos en cada coordenada, es decir, obliga a que la matriz A tenga m columnas de ceros,(el grafo correspondiente tendría m aristas a los que no se llega desde ninguno de las otras aristas, el grafo no es fuertemente conexo), y por tanto la matriz A es reducible.

Hay una contradicción, ya que sabemos que A es irreducible, y por tanto la hipótesis es falsa. Es decir, ningún rayo de R^+ se transforma mediante A en el elemento nulo.

(ii) No hay ningún rayo en el borde de R^+ que quede invariante por A. Lo razonamos por reducción al absurdo. Supongamos que existe un rayo del borde del cuadrante, $r = r[v]$, tal que $A(r) = r$; entonces, existe k, $0 < k < n$, tal que las primeras k coordenadas de v

[1]La demostración por reducción al absurdo es un tipo de argumento muy empleado en demostraciones matemáticas. Consiste en demostrar que una proposición matemática es verdadera, probando que si no lo fuera conduciría a una contradicción, por lo cual sería verdadera.

son cero (después de reordenarlas si fuese necesario); La condición $A(r) = r$ implica que A es de la forma

$$\begin{bmatrix} M & 0 \\ P & N \end{bmatrix}$$

donde M es $k \times k$, y N es de orden $(n-k) \times (n-k)$. Contradicción con que A es irreducible.

(iii) En este apartado vamos a utilizar el Teorema del punto fijo de Brouwer.

Para poder aplicar el Teorema de punto fijo de Brouwer a R^+, necesitamos comprender que este espacio se puede ver como un conjunto convexo, compacto y no vacío. Para ello, basta identificar cada rayo con su punto perteneciente al n-símplice

$$\{(x_1, x_2, \ldots, x_n) \ : \ x_1 + x_2 + \cdots + x_n = 1, \ x_i \geq 0, \ i = 0, \ldots, n\}.$$

En R^3 lo podemos visualizar. Cada rayo de R^+ se identifica mediante el punto de intersección del propio rayo con el triángulo T, siendo T la intersección del plano $x + y + z = 1$ con el octante positivo. Ver siguiente figura.

Hemos identificado R^+ con un n-símplice, por lo que cumple ser convexo, compacto y no vacío. Por tanto, como $A(R^+) \subset R^+$ por *(i)*, el Teorema del punto fijo de Brouwer asegura que hay un punto fijo de dicho triángulo, o equivalentemente, un rayo r invariante en R^+. Por *(ii)* sabemos que ningún rayo del borde del cuadrante es fijo, lo cual implica que r es un rayo positivo. Es decir, si tomamos un vector v en la dirección del rayo r (v es un autovector), v no tiene ninguna de sus coordenadas nula. Además, el autovalor asociado a v, λ, es positivo puesto que tanto v como su imagen pertenecen al cuadrante positivo, C^+.

Para acabar la demostración vamos a representar cada rayo por su punto que dista una unidad del origen de coordenadas. De este modo, hay una correspondencia biunívoca entre los rayos de R^+ y la esfera S^{n-1}. Y vamos a estudiarlo particularizando a $\mathbb{R}^3$, es análogo en $\mathbb{R}^n$.

En $\mathbb{R}^3$, para saber cómo actúa A al aplicársele sobre un plano, p, basta con saber cómo actúa sobre la circunferencia S^1_p que obtenemos de la intersección de S^2 con el plano p.

Supongamos que π es un plano invariante que contiene a r, sea L el arco $R^+ \cap S^1_\pi$, por pertenecer L al plano invariante y por *(i)*, se cumple que $A(L) \subset L$:

La aplicación A restringida a π, es una aplicación lineal y podemos encontrar su forma de Jordan. Sabemos que uno de sus autovalores es λ, por tanto, hay tres posibilidades:

(a) $\text{mult}_{\text{alg}}(\lambda) = \text{mult}_{\text{geom}}(\lambda) = 2$, en cuyo caso la forma de Jordan sería:

$$A|_\pi = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

(b) $\text{mult}_{\text{alg}}(\lambda) = 2$ y $\text{mult}_{\text{geom}}(\lambda) = 1$, en cuyo caso la forma de Jordan sería:

$$A|_\pi = \begin{bmatrix} \lambda & 0 \\ 1 & \lambda \end{bmatrix}$$

(c) $\text{mult}_{\text{alg}}(\lambda) = \text{mult}_{\text{geom}}(\lambda) = 1$, en cuyo caso la forma de Jordan sería:

$$A|_\pi = \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$$

Analizamos cada uno de los tres casos y vamos a llegar a la conclusión de que sólo es posible el caso (c).

(1) Vemos por reducción al absurdo que el caso (a) no es posible.

Supongamos que existen dos autovectores independientes asociados a λ. Tomando estos dos vectores como base del plano observamos que todos los vectores del plano son autovectores asociados al autovalor λ. Ya que suponiendo si v_1 y v_2 son los autovectores asociados a λ, cualquier vector w del plano se puede expresar como $w = a \cdot v_1 + b \cdot v_2$, y por ser A la matriz asociada a una transformación lineal, tenemos que la imagen de w sería $A \cdot w = a \cdot Av_1 + b \cdot Av_2 = a \cdot \lambda \cdot v_1 + b \cdot \lambda \cdot v_2 = \lambda \cdot a \cdot v_1 + b \cdot v_2 = \lambda \cdot w$.

Esto implica que A deja invariantes todos los rayos del plano, en particular, deja invariantes los rayos del borde de R^+, lo cuál contradice al apartado *(ii)*.

Por tanto, la hipótesis es incorrecta, no pueden existir dos autovectores independientes asociados a λ.

(2) Vemos ahora que el caso (b) no es posible.

Reducción al absurdo: Suponemos que la matriz $A|_\pi$ es semejante a la matriz del caso b).

Sabemos que dos matrices semejantes pueden pensarse como dos descripciones de una misma transformación lineal, pero con respecto a bases distintas. Vamos a analizar la transformación que conlleva la matriz de Jordan del caso b) y, por tanto, como estamos suponiendo que la matriz de $A|_\pi$ es semejante, va a realizar la misma transformación lineal, pero respecto de otra base, es decir, dejando fijos otros vectores.

Observamos que:

$$\begin{bmatrix} \lambda & 0 \\ 1 & \lambda \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a\lambda \\ a + b\lambda \end{bmatrix}$$

La aplicación lineal dada por

$$\begin{bmatrix} \lambda & 0 \\ 1 & \lambda \end{bmatrix}$$

deja fijos los rayos que pasan por $(0,1)$ y $(0,-1)$, y todos los demás rayos giran en sentido antihorario. Pero no todos los vectores sufren un giro de la misma amplitud, cuanto más cerca está el vector de uno de los vectores fijos menor es el giro, pues la imagen no puede sobrepasar al vector fijo.

Esto se comprueba fácilmente comparando el ángulo que forma el vector (a,b) con el semieje positivo OX, con el ángulo que forma su imagen $(a\lambda, a+b\lambda)$ con ese mismo semieje.

Si la matriz de $A|_\pi$ es semejante a esta matriz, va a realizar la misma transformación. Es decir, deja dos rayos fijos y, en este caso, todos los demás rayos giran en sentido antihorario sufriendo un giro de menor amplitud cuánto más cerca esté de los rayos fijos.

Vemos la contradicción:

- Los rayos fijos, a través de $A|_\pi$, no pueden ser del borde de R^+ (porque contradiría a (ii)).
- Los rayos fijos, a través de $A|_\pi$, no pueden ser distintos del eje vertical (porque si no el vector que pasa por $(0,1)$ sale de R^+, y esto contradice al apartado (i)).

Por tanto, la matriz de $A|_\pi$ no puede ser semejante a la del caso b).

Nota: Si en vez de tomar como forma de Jordan la matriz $\begin{bmatrix}\lambda & 0\\ 1 & \lambda\end{bmatrix}$ se toma la matriz $\begin{bmatrix}\lambda & 1\\ 0 & \lambda\end{bmatrix}$, la transformación que se conlleva es similar, la diferencia es que deja fijo el eje horizontal, y que los vectores giran en sentido horario. En $A|_\pi$ se llegaría a contradicción viendo que el semieje positivo OX, si es fijo contradice a (ii); y si no es fijo, hay rayos de R^+ que salen de r^+.

Por (1) deducimos que la multiplicidad geométrica de λ es 1, y por (2) deducimos que la multiplicidad algebraica de λ es también 1.**En conclusión, sólo se puede dar el caso (c)**

$$A|_\pi = \begin{bmatrix}\lambda & 0\\ 0 & \mu\end{bmatrix}.$$

Resumiendo, en este apartado hemos visto que el autovalor λ, correspondiente al rayo fijo, es simple (tiene multiplicidad algebraica 1), es positivo y tiene asociado un autovector estrictamente positivo.

(iv) Solo nos queda mostrar que $|\lambda| \geq |\mu|$ para el otro autovalor μ.

Por reducción al absurdo. Supongamos que el autovalor μ cumple que $|\lambda| < |\mu|$, distinguimos dos casos:

Caso 1: μ es un número real.

Sea v_μ el autovalor asociado a al autovalor μ tal que su rayo $r_\mu \notin R^+$ (porque si $r_\mu \in R^+$ ya estaría demostrado el teorema). Entonces A^n actúa en la circunferencia S^1 que representa todos los rayos del plano generado por v_λ y v_μ, y fija al conjunto de rayos $\{\pm r_\lambda, \pm r_\mu\}$.

Como $\lambda < |\mu|$, el movimiento que realiza A^n en S^1 tiene dos puntos de atracción, $\{\pm r_\mu\}$, y dos puntos de repulsión, $\{\pm r_\lambda\}$. Entonces r_μ atrae a uno de los dos puntos de $\partial R^+ \cap S^1$ fuera de R^+, lo cual no es posible.

Caso 2: μ es un número complejo.

Comenzamos recordando que, si μ es autovalor de A, entonces $\bar{\mu}$ también es autovalor de A. Además, si v es un autovector asociado a μ, se cumple que v es autovector asociado a $\bar{\mu}$.

Vamos a demostrar que, si P es el plano invariante correspondiente al autovalor μ, P no contiene ningún rayo de R^+:

En un principio se tiene tres posibilidades para $P \cap R^+$:

a) Que se corten en un único rayo.

b) Que parte de P esté contenido en R^+.

c) Que P no corte a R^+.

Vemos que a) es imposible:

Para que $P \cap R^+$ sea un rayo es necesario que dicho rayo pertenezca al borde de R^+. Entonces la imagen de ese rayo sería el mismo, ya que $A(R^+) \subset R^+$, el plano es invariante y su único elemento perteneciente a R^+ es este rayo. Esto es imposible porque contradice al apartado (ii) que dice que ningún rayo del borde de R^+ es invariante. En conclusión, $P \cap R^+$ no es un rayo.

Vemos que b) es imposible por reducción al absurdo:

Teniendo en cuenta que la circunferencia S^1 representa a todos los rayos de P, cada punto de la circunferencia representa al rayo que pasa por el origen y por ese punto, la colección de rayos de R^+ que está contenida en P se corresponde un arco L de dicha circunferencia.

Como A es una rotación en S^1 (porque es como actúa A en todo el espacio propio de μ), $A(L)$ no estaría contenida en L, lo cual contradice que $A(R^+) \subset R^+$. Por tanto, la opción b) también es imposible.

La conclusión es que sólo se puede dar la opción c), es decir, P no corta a R^+.

Demostramos que $|\lambda| \geq |\mu|$ por reducción al absurdo:

Supongamos que $|\lambda| < |\mu|$. Llamamos E^3 al espacio generado por P y por v_λ.

Por ser $|\lambda| < |\mu|$ la acción de A en E^3 muestra repulsión a la línea generada por v_λ, y una atracción al plano P. Esto implica que los rayos de $R^+ \backslash r_\lambda$ en E^3 se aproximan a P tanto como queramos si aplicamos repetidamente la aplicación de A, y por tanto nosotros concluimos que todos los rayos se salen de R^+ después de aplicar un número de veces, lo cual no es posible porque sabemos que $A(R^+) \subset R^+$. (Una demostración formal de esto es la siguiente: sea $r[x]$ un rayo de E^3 en $R^+ \backslash r_\lambda$ con $x = v_\lambda + w$, siendo $w \in P$ y $w \neq 0$. Entonces:

$$A^n\left(\frac{x}{|\mu|^n}\right) = \frac{A^n(v_\lambda) + A^n(w)}{|\mu|^n} = \frac{\lambda^n}{|\mu|^n} v_\lambda + \frac{\mu^n}{|\mu|^n} w.$$

Cuando $n \to \infty$ entonces $\frac{\lambda^n}{|\mu|^n} \to 0$, y por tanto el rayo $A^n(r[x])$ se acerca a P tanto como queramos, ya que el módulo de $\frac{\mu^n}{|\mu|^n} w \in P$ es constante).

Por tanto, la hipótesis es falsa. Podemos afirmar que $\lambda \geq |\mu|$. □

A.3. Uso de las APIs de Scopus

El término API es la abreviación en inglés de Application Programming Interfaces, que en español significa interfaz de programación de aplicaciones. Las APIs son interfaces web que permiten que las aplicaciones puedan comunicarse entre ellas a través de internet, por ejemplo:

1. Una página web puede interactuar con otra web.
2. Una aplicación de escritorio con un servidor.

El modelo de datos de Scopus está diseñado alrededor de la noción que los artículos están escritos por autores que a su vez están afiliados con instituciones. De una forma visual y simplificada, podemos representar el modelo de la siguiente forma, ver Figura (**??**).

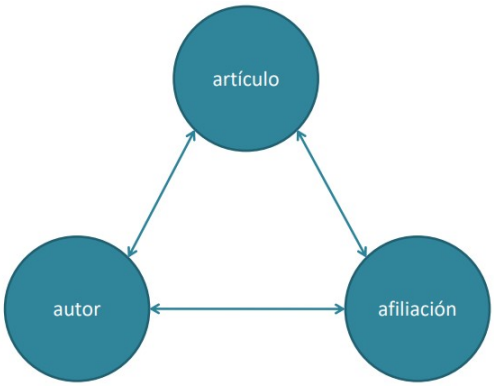

Figura A.2: Modelo de datos de Scopus simplificado.

Las APIs ofrecen las mismas funcionalidades que la interfaz de usuario de Scopus, pero en un formato legible también por máquinas que permite que el usuario o un software encuentre artículos, autores e instituciones en Scopus.

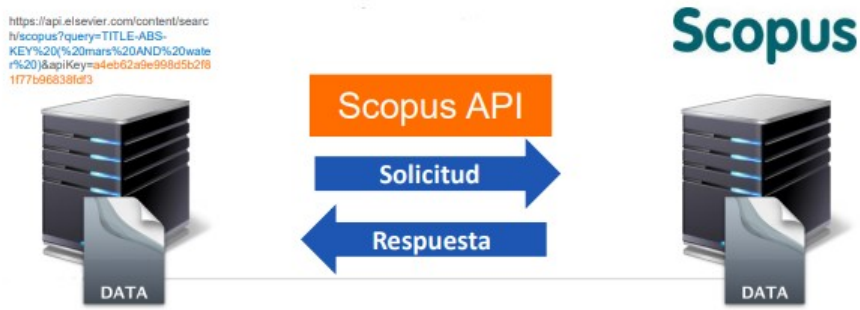

Figura A.3: Modelo solicitud respuesta de una API.

¿Cómo se genera una llamada a la API?

Para llamar a una API utilizaremos la UI de Scopus, lo que se recogerá de la API no es un archivo HTML que se muestra en un navegador, sino XML o JSON (otros formatos) que puede ser procesados por muchos programas. Las APIs normalmente no utilizan usuarios

o contraseñas, sino una APIKey y tokens de acceso.

Lo primero es conseguir una APIKey para tal fin nos registramos en la página web de Scopus, en la sección About Scopus buscamos el enlace Scopus API y solicitamos una APIKey la cual nos dará acceso a diversas APIs de Elsevier (Scopus). Para obtener la funcionalidad completa de las APIs, se necesitará también una APIKey adicional, llamado token institucional (insttoken) la cual tiene ciertas restricciones, por ejemplo, el de ser revocado en cualquier momento, sin previo aviso. Esta se solicita a través de un correo electrónico al área de soporte de Scopus, mencionando nuestra clave API y enviando la solicitud desde nuestro correo electrónico institucional como prueba de ser estudiante, personal, profesor, etc.

Debemos tener en cuenta que el token institucional estará vinculado a la clave API, además es de uso personal e intransferible, por lo que en el presente trabajo no se mostrarán las APIKeys utilizadas. Por último, es recomendable leer detenidamente la documentación y Las Políticas de Uso.

APIs Interactivas

En la sección Scopus API vamos al enlace Interactive APIs y ya podemos hacer uso de las APIs de Scopus con Scopus Interactive APIs, o de las APIs de SciVal con SciVal Interactive APIs.

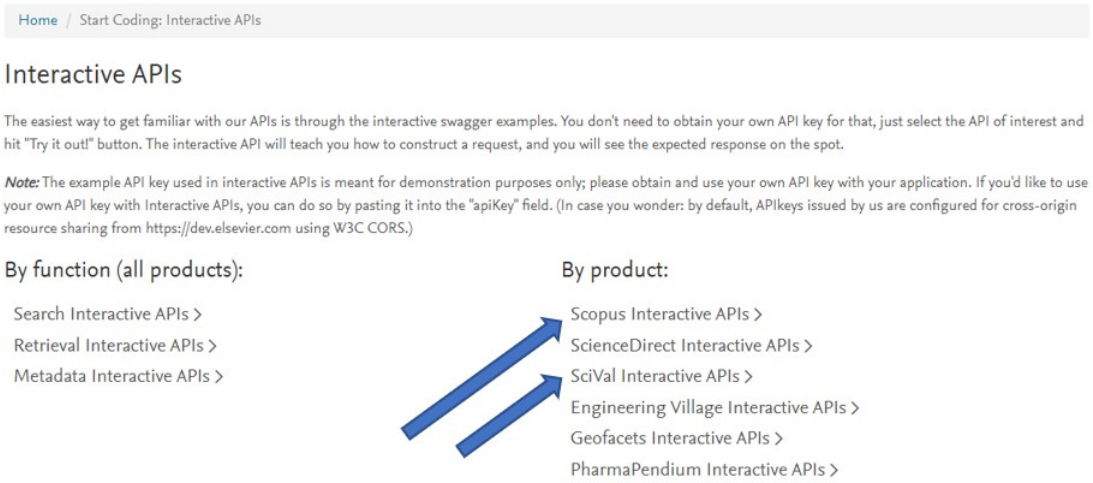

Figura A.4: Selección de las APIs de Elsevier. Web de Scopus.

Una vez dentro de ambos enlaces vamos a utilizar las siguientes opciones:

- Abstract Citation Count API, perteneciente a Scopus
- Citations Overview API, perteneciente a Scopus
- SciVal Institution Lookup API, perteneciente a SciVal

Vamos a ejemplificar la situación de artículos, a pesar de ser una extracción sistematizada se tiene la limitación de obtener 200 resultados por consulta, por lo que si el resultado es mayor, debemos consultar repetidas veces. Para llamar una API debemos rellenar el siguiente tipo de formulario, ver Figura (A.4).

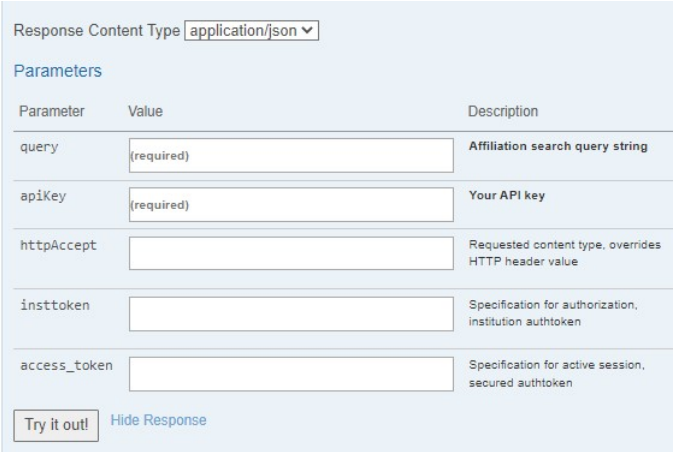

Figura A.5: Campos requeridos para llamar a una API. Web de Scopus.

- Note que previo a rellenar los campos el tipo de archivo que se recogerá es de tipo JSON.
- En el campo query: affilcountry(Peru) and affil(Revista Peruana de Medicina Experimental y Salud) and pubyearaft 2017 and pubyearbef 2021 and (doctype(ar)) field=eid,title,citedby-count and sort=-citedbycount and count=200
- Los comandos están en inglés, y son intuitivos, pero se recomienda leer la documentación para su manejo, el último comando pide obtener el máximo de resultados posibles por consulta, sabemos que el resultado total es mayor a 200, por lo que la siguiente llamada se debe modificar affilcountry(Peru) and affil(Revista Peruana de Medicina Experimental y Salud) and pubyearaft 2017 and pubyearbef 2021 and (doctype(ar)) field=eid,title,citedby-count and sort=-citedbycount and count=200&start=200, donde claramente piden contar la cantidad máxima posible por consulta pero ahora se pide que empiece desde 200.
- El campo apiKey es obligatorio, pero como se mencionó muchas veces no es suficiente para muchas funcionalidades.
- El campo insttoken es el token institucional.
- Al darle clic a: Try it out! se desplega unos cuadros con la información deseada en forma de link, los mostramos parcialmente ya que dicho link viene concatenada con nuestra APIKey.

Figura A.6: Resultado de la llamada de la API. Web de Scopus.

Al acceder al link tenemos el archivo xml/JSON, el cual puede ser exportado a un archivo Excel.

Como una de las finalidades de los archivos JSON es serializar y transmitir datos estructurados, además de incluso tener una lectura sencilla; una vez exportados los datos del archivo JSON estos llegan a Excel con múltiples filtros debido a la sobreinformación recogida.

Consideraciones finales

Concytec puede acceder a una base de 4 mil registros como máximo, mediante la API, siempre y cuando se cuente con la "CLAVE INSTITUCIONAL" (Insttoken). Ya que, con el clave (key) personal como usuario del Concytec, la API no permite acceder o extraer a la información de la base de datos de Scopus.

En ese sentido, para poder extraer los registros por autor se debe solicitar internamente a la oficina que tiende los derechos de la clave institucional, a fin de acceder a los registros de la base de datos de Scopus y de esta manera construir la matriz que requiere el Algoritmo para construir la información que se utilizará para clasificar artículos, revistas y autores.

Concytec posee licencias con Scopus que se renuevan anualmente, esta licencia le permite acceder a una clave o claves institucionales, dependiente las cláusulas del contrato. Por lo que, es importante solicitar dicha clave para consumir la información que requiere el algoritmo.

A.4. Primer Algoritmo utilizando Python

Listing A.1: Primer Algoritmo de Prueba

```
import numpy as np
from fractions import Fraction
##############################
### PRIMER CODIGO PARA Analisis de la Produccion cientfica utilizando
 algoritmos matemticos
###  CONCYTEC
##############################
my_dp = Fraction(1,4)
Mat = np.matrix([[0,0,Fraction(1,2),Fraction(1,2)],
[Fraction(1,3),0,0,0],[Fraction(1,3),0,0,Fraction(1,2)],[Fraction(1,3),1,
Fraction(1,2),0]])
Ex = np.ones((4,4))
Ex[:] = my_dp
print(Ex)
beta = 0.85
M = beta * Mat + ((1-beta) * Ex)
m=np.dot(np.transpose(M),M)


A=np.zeros((50, 4))
B=np.zeros((50,1))
E=np.zeros((50,1))

A[0]=np.array([np.diag(np.transpose(M)*M)/np.linalg.norm(np.diag
(np.transpose(M)*M))])
x_0=np.transpose(np.array([np.diag(np.transpose(M)*M)
/np.linalg.norm(np.diag(np.transpose(M)*M))]))
print("Vector inicial de la iteracin")
print(np.transpose(x_0))

for i in range(1,50):
    A[i]=np.dot(m,np.transpose(A[i-1]))/np.linalg.norm(
    np.dot(m,np.transpose(A[i-1])))
    B[i]=np.dot(np.transpose(A[i]),A[i-1])
    #print(B[i])
    E[i]=(B[i]-B[i-1])/B[i-1]
    print("-------------------------")
    print("Iteracin nmero")
    print(i)
    print("Vector de importancia")
    print(A[i])
    print("Error relativo")
    print(E[i])
    if (E[i] < 10**(-20)):
        print("Error relativo alcanzado")
        print ("se necesitaron", i, "iteraciones")
        break
```

Por lo tanto, para el caso de cuatro artículos el algoritmo arroja los siguientes resultados:

Listing A.2: Resultados del Primer Algoritmo de Prueba

```
Iteracin nmero
1
```

```
Vector de importancia
[0.37853746 0.70309643 0.51909324 0.30480652]
Error relativo
[inf]
-------------------------
Iteraci n  n mero
2
Vector de importancia
[0.37875597 0.70751544 0.51923888 0.29386527]
Error relativo
[0.01649363]
-------------------------
Iteraci n  n mero
3
Vector de importancia
[0.37823915 0.70963793 0.51861391 0.29049745]
Error relativo
[6.14056369e-05]
-------------------------
Iteraci n  n mero
4
Vector de importancia
[0.37803344 0.71041055 0.51837874 0.28929404]
Error relativo
[7.18109694e-06]
-------------------------
Iteraci n  n mero
5
Vector de importancia
[0.37795788 0.71068881 0.51829422 0.28886045]
Error relativo
[9.32242444e-07]
-------------------------
Iteraci n  n mero
6
Vector de importancia
[0.37793048 0.71078903 0.51826382 0.28870419]
Error relativo
[1.21068806e-07]
-------------------------
Iteraci n  n mero
7
Vector de importancia
[0.37792059 0.71082515 0.51825288 0.28864787]
Error relativo
[1.57231583e-08]
-------------------------
Iteraci n  n mero
8
Vector de importancia
[0.37791702 0.71083816 0.51824893 0.28862757]
Error relativo
[2.04196294e-09]
-------------------------
Iteraci n  n mero
9
Vector de importancia
[0.37791574 0.71084285 0.51824751 0.28862026]
```

```
Error relativo
[2.65189426e-10]
-------------------------
Iteraci n  n mero
10
Vector de importancia
[0.37791527 0.71084454 0.518247   0.28861762]
Error relativo
[3.44398954e-11]
-------------------------
Iteraci n  n mero
11
Vector de importancia
[0.37791511 0.71084515 0.51824682 0.28861667]
Error relativo
[4.47264448e-12]
-------------------------
Iteraci n  n mero
12
Vector de importancia
[0.37791505 0.71084537 0.51824675 0.28861633]
Error relativo
[5.80979709e-13]
-------------------------
Iteraci n  n mero
13
Vector de importancia
[0.37791502 0.71084545 0.51824672 0.28861621]
Error relativo
[7.54951657e-14]
-------------------------
Iteraci n  n mero
14
Vector de importancia
[0.37791502 0.71084548 0.51824672 0.28861616]
Error relativo
[9.54791801e-15]
-------------------------
Iteraci n  n mero
15
Vector de importancia
[0.37791501 0.71084549 0.51824671 0.28861615]
Error relativo
[1.33226763e-15]
-------------------------
Iteraci n  n mero
16
Vector de importancia
[0.37791501 0.71084549 0.51824671 0.28861614]
Error relativo
[2.22044605e-16]
-------------------------
Iteraci n  n mero
17
Vector de importancia
[0.37791501 0.71084549 0.51824671 0.28861614]
Error relativo
[-2.22044605e-16]
```

```
Error relativo alcanzado
se necesitaron 17 iteraciones
```

A.5. Código ordenación artículos utilizando Python

Listing A.3: Código para la ordenación de artículos

```
import numpy as np
from fractions import Fraction
import pandas as pd

############################
### CDIGO PARA An lisis de la Producci n cient fica utilizando
### algoritmos matem ticos
###          CONCYTEC
############################

#Importando los datos
test_csv = 'matrizarticulos.csv'
matriz = np.matrix(pd.read_csv(test_csv, sep=';', header=None)).
reshape((373,373))

#Preparando la matriz modelo
numiter = np.matrix(matriz).shape[0] #dimension de matriz base
Matriz = np.empty((numiter,0)) #futura matriz estoc stica
my_dp = Fraction(1,numiter) #coeficientes necesarios

### preparando matriz estoc stica
### Iteraci n que consigue matriz estoc stica
SumCol=np.einsum('ij->j',matriz)
#print(SumCol)
#print(SumCol[1])
for i in range(0,numiter):
    if SumCol[i] != 0:
        Matriz=np.append(Matriz,np.reshape((1/SumCol[i])*matriz[i],(numiter,1))
        , axis=1)
    else:
        Matriz=np.append(Matriz,np.reshape(matriz[i],(numiter,1)), axis=1)

######### Hallando matriz irreducible
Ex = np.ones((numiter,numiter)) #matriz necesaria para el modelo
Ex[:] = my_dp
beta = 0.85
#Haciendo densa a la matriz del modelo
M = beta * Matriz + ((1-beta) * Ex)

# Preparando matrices para el proceso iterativo
# del M todo de las Potencias

m=np.dot(np.transpose(M),M)
A=np.zeros((1000, numiter))
B=np.zeros((1000,1))
E=np.zeros((1000,1))

A[0]=np.array([np.diag(np.transpose(M)*M)/np.linalg.norm(np.diag(np.transpose(
M)*M))])
x_0=np.transpose(np.array([np.diag(np.transpose(M)*M)/np.linalg.no
rm(np.diag(np.transpose(M)*M))]))
print("Vector inicial de la iteraci n")
print(np.transpose(x_0))
```

```
#Iniciando proceso iterativo:
for i in range(1,1000):
    A[i]=np.dot(m,np.transpose(A[i-1]))/np.linalg.norm(np.dot(m,np
    .transpose(A[i-1])))
    B[i]=np.dot(np.transpose(A[i]),A[i-1])
    E[i]=(B[i]-B[i-1])/B[i-1]
    print("-------------------------")
    print("Iteraci n n mero:", i)
    print("Error relativo:",E[i])
    print(np.where(max(A[i]) == A[i]))
    if (E[i] < 10**(-20)):
        print("Error relativo alcanzado")
        print ("se necesitaron", i, "iteraciones")
#         print(np.where(max(A[i]) == A[i])) #Posici n de mayor importancia
# Notar que por el modo de escritura de Python se le debe sumar 1.
        print("Vector de ordenamiento de menor a mayor:")
        print(np.argsort(A[i])+np.ones((1,numiter),int))
        break
```

Resultados

Los resultados son los siguientes, ponemos las primeras y últimas iteraciones:

Listing A.4: Resultados de ordenación de artículos

```
-------------------------
Iteraci n n mero:  1
Error relativo:  [inf]
(array([34], dtype=int64),)
-------------------------
Iteraci n n mero:  2
Error relativo:  [0.05533004]
(array([34], dtype=int64),)
-------------------------
Iteraci n n mero:  3
Error relativo:  [0.01213175]
(array([34], dtype=int64),)
.
.
.
.
.
.
.
Iteraci n n mero:  190
Error relativo:  [2.22044605e-16]
(array([348], dtype=int64),)
-------------------------
Iteraci n n mero:  191
Error relativo:  [3.33066907e-16]
(array([348], dtype=int64),)
-------------------------
Iteraci n n mero:  192
Error relativo:  [2.22044605e-16]
(array([348], dtype=int64),)
-------------------------
Iteraci n n mero:  193
```

```
Error relativo:  [1.11022302e-16]
(array([348], dtype=int64),)
-------------------------
Iteraci n n mero:  194
Error relativo:  [2.22044605e-16]
(array([348], dtype=int64),)
-------------------------
Iteraci n n mero:  195
Error relativo:  [-1.11022302e-16]
(array([348], dtype=int64),)
Error relativo alcanzado
se necesitaron 195 iteraciones
Vector de ordenamiento de menor a mayor:
[[370 369 368 367   1  ...      ...      ...  11 358 169 ...
... ... 304 346  61 260 193 324 328  35 251 262 248 349]]
```

A.6. Código ordenación revistas utilizando Python

Listing A.5: Código para la ordenación de artículos

```
import numpy as np
from fractions import Fraction
import pandas as pd

############################
### CDIGO PARA An lisis de la Producci n cient fica utilizando
### algoritmos matem ticos
###          CONCYTEC
############################

#Importando los datos
test_csv = 'matrizrevistas.csv'
matriz = np.matrix(pd.read_csv(test_csv, sep=';', header=None)).reshape((15,15))

#Preparando la matriz modelo
numiter = np.matrix(matriz).shape[0] #dimension de matriz base
Matriz = np.empty((numiter,0)) #futura matriz estoc stica
my_dp = Fraction(1,numiter) #coeficientes necesarios

### preparando matriz estoc stica
### Iteraci n que consigue matriz estoc stica
SumCol=np.einsum('ij->j',matriz)
#print(SumCol)
#print(SumCol[1])
for i in range(0,numiter):
    if SumCol[i] != 0:
        Matriz=np.append(Matriz,np.reshape((1/SumCol[i])*matriz[i],
        (numiter,1)), axis=1)
    else:
        Matriz=np.append(Matriz,np.reshape(matriz[i],(numiter,1)), axis=1)

######### Hallando matriz irreducible
Ex = np.ones((numiter,numiter)) #matriz necesaria para el modelo
Ex[:] = my_dp
beta = 0.85
#Haciendo densa a la matriz del modelo
M = beta * Matriz + ((1-beta) * Ex)

# Preparando matrices para el proceso iterativo
# del M todo de las Potencias

m=np.dot(np.transpose(M),M)
A=np.zeros((1000, numiter))
B=np.zeros((1000,1))
E=np.zeros((1000,1))

A[0]=np.array([np.diag(np.transpose(M)*M)/np.linalg.norm(np.
diag(np.transpose(M)*M))])
x_0=np.transpose(np.array([np.diag(np.transpose(M)*M)/np.
linalg.norm(np.diag(np.transpose(M)*M))]))
print("Vector inicial de la iteraci n")
print(np.transpose(x_0))
```

```
#Iniciando proceso iterativo:
for i in range(1,1000):
    A[i]=np.dot(m,np.transpose(A[i-1]))/np.linalg.norm(np.dot(m,np.
    transpose(A[i-1])))
    B[i]=np.dot(np.transpose(A[i]),A[i-1])
    E[i]=(B[i]-B[i-1])/B[i-1]
    print("-------------------------")
    print("Iteraci n n mero: ",i)
    print("Error relativo: ",E[i])
    print(np.where(max(A[i]) == A[i]))
    if (E[i] < 10**(-20)):
        print("Error relativo alcanzado")
        print ("se necesitaron", i, "iteraciones")
#         print(np.where(max(A[i]) == A[i])) #Posici n de mayor importancia
# Notar que por el modo de escritura de Python se le debe sumar 1.
        print("Vector de ordenamiento de menor a mayor:")
        print(np.argsort(A[i])+np.ones((1,numiter),int))
        break
```

Resultados

Los resultados son los siguientes, note que al tener menos información el truncamiento del proceso se da rapidamente.

Listing A.6: Resultados de ordenación de revistas

```
Vector inicial de la iteraci n
[[0.00664005 0.00461342 0.02698161 0.01654491 0.00371797 0.00460301
  0.0038156  0.00468759 0.01765609 0.01622206 0.99898821 0.00950667
  0.0022831  0.00543165 0.01356118]]
-------------------------
Iteraci n n mero:  1
Error relativo:  [inf]
(array([10], dtype=int64),)
-------------------------
Iteraci n n mero:  2
Error relativo:  [0.01416531]
(array([10], dtype=int64),)
-------------------------
Iteraci n n mero:  3
Error relativo:  [8.31047817e-06]
(array([10], dtype=int64),)
-------------------------
Iteraci n n mero:  4
Error relativo:  [2.26278671e-08]
(array([10], dtype=int64),)
-------------------------
Iteraci n n mero:  5
Error relativo:  [6.27502494e-11]
(array([10], dtype=int64),)
-------------------------
Iteraci n n mero:  6
Error relativo:  [1.74193993e-13]
(array([10], dtype=int64),)
-------------------------
Iteraci n n mero:  7
Error relativo:  [3.33066907e-16]
```

```
(array([10], dtype=int64),)
-------------------------
Iteraci n n mero: 8
Error relativo: [1.11022302e-16]
(array([10], dtype=int64),)
-------------------------
Iteraci n n mero: 9
Error relativo: [-1.11022302e-16]
(array([10], dtype=int64),)
Error relativo alcanzado
se necesitaron 9 iteraciones
Vector de ordenamiento de menor a mayor:
[[ 7  9  5 13 14  8  6  2  1  4 12  3 15 10 11]]
```

A.7. Código ordenación autores utilizando Python

Listing A.7: Código para la ordenación de artículos

```
import numpy as np
from fractions import Fraction
import pandas as pd

############################
### CDIGO PARA An lisis de la Producci n cient fica utilizando
### algoritmos matem ticos
###          CONCYTEC
############################

#Importando los datos
test_csv = 'matrizautores.csv'
matriz = np.matrix(pd.read_csv(test_csv, sep=';', header=None)).reshape((36,36))

#Preparando la matriz modelo
numiter = np.matrix(matriz).shape[0] #dimension de matriz base
Matriz = np.empty((numiter,0)) #futura matriz estoc stica
my_dp = Fraction(1,numiter) #coeficientes necesarios

### preparando matriz estoc stica
### Iteraci n que consigue matriz estoc stica
SumCol=np.einsum('ij->j',matriz)
#print(SumCol)
#print(SumCol[1])
for i in range(0,numiter):
    if SumCol[i] != 0:
        Matriz=np.append(Matriz,np.reshape((1/SumCol[i])*matriz[i],
        (numiter,1)), axis=1)
    else:
        Matriz=np.append(Matriz,np.reshape(matriz[i],(numiter,1)), axis=1)

######### Hallando matriz irreducible
Ex = np.ones((numiter,numiter)) #matriz necesaria para el modelo
Ex[:] = my_dp
beta = 0.85
#Haciendo densa a la matriz del modelo
M = beta * Matriz + ((1-beta) * Ex)

# Preparando matrices para el proceso iterativo
# del M todo de las Potencias

m=np.dot(np.transpose(M),M)
A=np.zeros((1000, numiter))
B=np.zeros((1000,1))
E=np.zeros((1000,1))

A[0]=np.array([np.diag(np.transpose(M)*M)/np.linalg.norm(np.
diag(np.transpose(M)*M))])
x_0=np.transpose(np.array([np.diag(np.transpose(M)*M)/np.
linalg.norm(np.diag(np.transpose(M)*M))]))
print("Vector inicial de la iteraci n")
print(np.transpose(x_0))
```

```
#Iniciando proceso iterativo:
for i in range(1,1000):
    A[i]=np.dot(m,np.transpose(A[i-1]))/np.linalg.norm(np.dot(m,np.
    transpose(A[i-1])))
    B[i]=np.dot(np.transpose(A[i]),A[i-1])
    E[i]=(B[i]-B[i-1])/B[i-1]
    print("-------------------------")
    print("Iteraci n n mero:", i)
    print("Error relativo:",E[i])
    print(np.where(max(A[i]) == A[i]))
    if (E[i] < 10**(-20)):
        print("Error relativo alcanzado")
        print ("se necesitaron", i, "iteraciones")
#         print(np.where(max(A[i]) == A[i])) #Posici n de mayor importancia
# Notar que por el modo de escritura de Python se le debe sumar 1.
        print("Vector de ordenamiento de menor a mayor:")
        print(np.argsort(A[i])+np.ones((1,numiter),int))
        break
```

Resultados

Los resultados son los siguientes, mostramos las primeras y últimas iteraciones.

Listing A.8: Resultados de ordenación de revistas

```
Vector inicial de la iteraci n
[[2.96509521e-04 1.17636792e-02 3.20001150e-02 6.27503099e-02
  5.62838608e-02 6.40445437e-03 2.48014555e-02 5.88353959e-02
  2.99474616e-02 4.08467482e-02 3.69523715e-02 7.04098921e-01
  1.10062172e-01 2.81315929e-02 7.84655426e-02 6.18401270e-03
  8.40110309e-03 1.75179472e-01 8.83048816e-03 6.65864636e-02
  7.02095947e-02 2.20899430e-02 4.55636297e-02 9.51411941e-03
  6.05867788e-02 2.38411050e-02 5.28786203e-02 2.65685982e-02
  3.20427378e-02 2.88468183e-02 5.47474608e-02 5.83513584e-01
  1.81045621e-02 1.73639270e-01 1.49232315e-01 1.41915104e-01]]
-------------------------
Iteraci n n mero: 1
Error relativo: [inf]
(array([11], dtype=int64),)
-------------------------
Iteraci n n mero: 2
Error relativo: [0.07314025]
(array([31], dtype=int64),)
-------------------------
Iteraci n n mero: 3
Error relativo: [0.00383615]
(array([31], dtype=int64),)
-------------------------
Iteraci n n mero: 4
Error relativo: [0.00073881]
(array([31], dtype=int64),)
-------------------------
Iteraci n n mero: 5
Error relativo: [0.00014477]
(array([31], dtype=int64),)
-------------------------
Iteraci n n mero: 6
Error relativo: [2.82798509e-05]
```

```
(array([31], dtype=int64),)
-------------------------
.
.
.
.
-------------------------
Iteraci n n mero:  21
Error relativo:  [7.77156117e-16]
(array([31], dtype=int64),)
-------------------------
Iteraci n n mero:  22
Error relativo:  [1.11022302e-16]
(array([31], dtype=int64),)
-------------------------
Iteraci n n mero:  23
Error relativo:  [0.]
(array([31], dtype=int64),)
Error relativo alcanzado
se necesitaron 23 iteraciones
Vector de ordenamiento de menor a mayor:
[[ 1 24  6  2 17 16  3 19 14 26 33 22  9 28 11 10  7 29 30  4 27  8 15 13
  21 23 25 31 20  5 18 36 34 35 12 32]]
```

Bibliografía

[1] Dieter Jungnickel, *Graphs, Networks and Algorithms*,Springer,2007

[2] R. Criado, M. Romance y L. Solá, *Teoría de Perron-Frobenius: importancia, poder y centralidad*, La Gaceta de la RSME, 2014

[3] C. Chávez Vega, *Álgebra Lineal*, UNI, 2004

[4] J. Rojas Tenazoa, *Aproximación matemática y computacional del motor de búsqueda Google*, Universidad Nacional De La Amazonía Peruana, 2016

[5] A. Inés Armas, *El teorema de Perron-Frobenius y su aplicación en el algoritmo de búsqueda de Google*, Universidad de la Rioja, 2015

[6] L. Miralles Millas, *Teorema de Perron-Frobenius Aplicación en el JCR (Eigenfactor TM Score en el JCR)*, Universidad de Zaragoza, 2019

[7] M. Baena Marzo, *Una demostración geométrica del teorema de Perron-Frobenius*, Universidad De Educación A Distancia, 2015

[8] E. Quispe Calderón, *Prueba geométrica del Teorema de Perron-Frobenius y su aplicación Al Google*, Universidad Nacional Del Altiplano, 2017

[9] Fan Chung, *A brief survey of PageRank algorithms*, University of California, San Diego.

[10] Moya Anegón, F., Herrán Páez, E., Bustos González, A., Vargas Quesada, B., Guzmán Morales, D., Corera-Álvarez, E., ... Dinu, R. (2019). *Principales indicadores bibliométricos de la actividad científica peruana. 2012-2017.*

[11] Ilse Ipsen, Rebecca M. Wills, *Analysis and Computation of Googles PageRank*, North Carolina State University

[12] L. Zack, R. Lamb, S. Ball, *An application of Google PageRank to NFL rankings*, Involve a journal of mathematics, 2012.

[13] P. Fernández Gallardo, *El secreto de Google y el Álgebra Lineal*, Universidad Autónoma de Madrid,Departamento de Matemáticas.

[14] ELSEVIER, *Research Metrics Guidebook - Research Intelligence*, 2019.