Informe Trabajo Práctico 2

Métodos Númericos 2C 2023

Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Grupo 6

Desarrollo y análisis de redes a partir del trabajo con autovalores y autovectores.

Integrante
Levi, Cindy
Zeitoune, Maia Sol
Azar, Agustin

1. Introducción

En este trabajo realizaremos análisis sobre redes a partir de su interpretación como matrices utilizando autovectores y autovalores.

1.1. Autovalores y autovectores

Dada una matriz $A \in \mathbb{R}^{n \times n}$ y un vector $v \in \mathbb{R}^n$, v es autovector de A si se satisface:

$$Av = \lambda v$$

para algún escalar λ.
Decimos entonces que v es un autovector de A con λ su autovalor asociado.

1.2. Método de la potencia y deflación

El método de la potencia es un método para calcular los autovalores de una matriz.
Sea $A \in \mathbb{R}^{n \times n}$, $\lambda^1, \lambda^2, ..., \lambda^n$ sus n autovalores y $v^1, v^2, ..., v^n$ los autovectores asociados, sea $q^0 \in \mathbb{R}^n$ tal que $||q^0||_2 = 1$, la sucesión $\{q^k\}$ se define como:

$$\begin{aligned} &\text{Para k} = 1 \ ... \\ &z^k = Aq^{k-1} \\ &q^k = \frac{z^k}{||z^k||_2} \end{aligned} \tag{1}$$

Esta su sesión converge al autovector v^1 y además $\lambda^k = (q^k)^t A q^k$ converge a λ^1.
El método de deflación consiste en obtener todos los autovalores de una matriz $A \in \mathbb{R}^{n \times n}$ utilizando el método de la potencia. sea λ_1 el autovalor obtenido con el método de la potencia y v_1 un autovector asociado, la matriz $B = \lambda_1 v_1^t v_1$ tiene los autovalores $0, \lambda_2, ..., \lambda_n$ con autovectores asociados $v_1, ..., v_n$, estos son los mismos que los de la matriz A original.
Si volvemos a aplicar el método de la potencia en esta matriz B obtendremos el autovector λ_2, ya que λ_1 no es autovalor, y su autovector asociado v_2.

1.3. Matriz de adyacencia

Un grafo es una estructura abstracta que consta de dos conjuntos, un conjunto de vértices denominado V y un conjunto de aristas E, donde una arista es una tupla de dos nodos, ej: $e_1 = (v, w) : v, w \in V$ donde una arista representa la existencia de una relación entre esos dos nodos.
Entre las formas de representar los grafos se encuentra la matriz de adyacencia. Se dice matriz de adyacencia del grafo $G = (V, E)$ a la matriz $A \in \mathbb{R}^{n \times n}$ donde $n = \#V$ y se cumple que:

$$A_{ij} = \begin{cases} 1 & \text{si } (i,j) \in E \\ 0 & cc \end{cases}$$

Notemos que esta matriz es simétrica si el grafo es no dirigido.

1.4. Matriz Laplaciana

La matriz laplaciana es una matriz construida a partir de la matriz de adyacencia de una red y su descomposición en autovectores permite analizar la estructura de una red.
Se construye con la formula $L = D - A$ donde A es la matriz de adyacencia y D es una matriz diagonal definida como:

$$D_{ii} = \sum_{j=1}^{n} A_{ij}$$

Es decir, los valores de D indican la cantidad de vecinos de un nodo.

1.5. Centralidad

La centralidad de una red es una medida que se puede utilizar para establecer un criterio de ordenamiento para los nodos de la red en base a su influencia o importancia en la misma.
Existen varios criterios para está medida, como por ejemplo basarse en el grado de un nodo.
En este informe trabajaremos sobre la centralidad del autovector donde el valor de centralidad de un nodo también considera la centralidad de sus vecinos. La centralidad del autovector se puede escribir de la siguiente forma:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j$$

Que es equivalente a la ecuación del autovector

$$\lambda x = Ax$$

2. Método de la potencia con deflación

2.1. Desarrollo

Para llevar a cabo experimentos relacionados a las redes a partir de los autovectores y autovalores de las matrices dadas, realizamos una implementación del método de la potencia con deflación en c++ explicado en la sección 1.2. utilizando la biblioteca eigen. En primer lugar, definimos el método de la potencia como un método que recibe la matriz, un épsilon y la cantidad de iteraciones máxima. Como la sucesión q^k definida en la introducción, en caso de converger, lo hace al autovector dominante, definimos una lista de números aleatorios y aplicamos la sucesión q^k tantas veces como haya sido indicado por parámetro. Se espera que luego de esa cantidad de iteraciones se alcance un resultado aproximado del valor real del autovector. Para evitar tener que realizar siempre la cantidad máxima de iteraciones, se calcula la diferencia entre el autovector resultante en la iteración i-1 y la iteración i, en caso de que la diferencia sea menor a épsilon, se decide que la diferencia es lo suficientemente chica como para desestimarla y devolvemos el autovector encontrado. Luego, faltaría encontrar el autovalor asociado al autovector encontrado. Para ello, sabemos que:

$$Av = \lambda v \;\Rightarrow\; v^t A v = \lambda v^t v \;\Rightarrow\; \frac{v^t A v}{v^t v} = \lambda$$

Podemos ver que para encontrar el autovalor asociado al autovector debemos resolver $\frac{v^t A v}{v^t v}$. A continuación, el pseudocódigo del método de la potencia descripto:

```
metodoDeLaPotencia(matriz, eps, itMax){
      autovec = createRandomArray(size = matriz.length)
      prev_autoval = 0;
      autoval = 1e6;
      for(int i = 0; i < itMax && abs(diff-prev_diff) > eps ; ++i){
          prev_autoval = autoval;
          vector_resultado = matriz * autovec;
          autovec = vector_resultado.normalized();
          autoval = autovec.T * matriz * autovec;
      }
      return autoval, v;
  }
```

Luego, queremos obtener el siguiente autovalor de la matriz. Para ello, podemos pensar que si obtenemos una matriz con los mismos autovalores pero sin el autovalor dominante, podremos obtener el próximo autovalor dominante λ_2. De esta forma, redefinimos la matriz de la siguiente forma:

$$A_2 = A_1 - \lambda_1 v_1 v_1^t$$

Si a A_2 multiplicamos por derecha a v_1, entonces el resultado es nulo. En cambio si a A_2 multiplicamos por derecha el autovector v_i de A_1, el resultado sera $\lambda_i v_i$, lo que indica que ese autovector es también autovector de A_2. Por lo tanto, escribimos la siguiente función que dada una matriz, la cantidad de autovalores buscados, el épsilon y la cantidad de iteraciones máxima, por cada autovalor buscado llama al método de la potencia, obtiene dicho autovalor junto al autovector asociado y luego genera la matriz A_2 para obtener el próximo autovalor.

```
void metodoPotenciaConDeflacion(matriz,  cant_autoval, eps, itMax){
    for( int i = 0; i < cant_autoval ; ++i){
        autovalor, autovector = metodoDeLaPotencia(matriz, eps, itMax);
        autoVal[i] = autovalor;
        autoVec.row(i) = autovector;
        matriz = matriz - autovalor * autovector * autovector.T;
    }
}
```

2.2. Experimentación

Para comprobar que nuestra implementación del método de la potencia funcione realizamos pruebas analizando la correctitud de los autovalores y los autovectores que calcula nuestra implementación. Utilizamos el método de construcción de una matriz A partir de una matriz D diagonal que contiene los autovalores conocidos y una matriz de Householder H, tal que $H^t D H = A$, por lo que los valores contenidos en la diagonal de D serán efectivamente los autovalores de A por ser matrices semejantes.
Para ver la correctitud de nuestros autovectores hicimos pruebas sobre matrices con dimensiones desde 2x2 a 100x100, luego calculamos los autovectores con utilizando 20.000 y 50.000 iteraciones para después medir el error cuadrático medio entre el vector de autovectores conocidos y el vector de autovalores calculado.

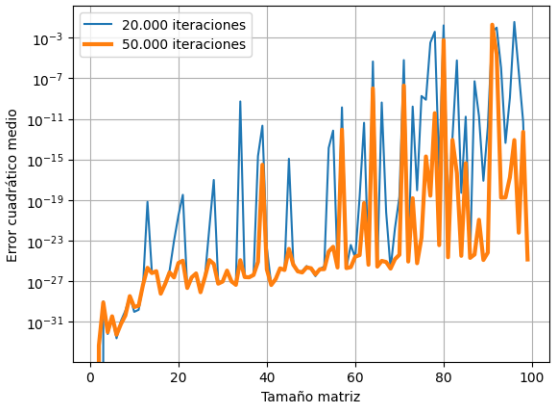


Figura 1: Comparación del error cuadrático medio con los autovalores conocidos

Podemos ver como en la figura 1 el ECM siempre es menor a 10^{-2} por lo que podemos concluir que nos brinda valores confiables.

También quisimos verificar que efectivamente los autovectores generados por nuestra función sean correctos, por lo que generamos de forma constructiva matrices A con autovalores conocidos de tamaño entre 2 y 100, y luego para cada una de ellas calculamos el error cuadrático medio de la diferencia entre Av_1 y $\lambda_1 v_1$, siendo v_1 el autovector dominante de la matriz A y λ_1 el autovalor correspondiente. Realizamos esto tanto con nuestra función como con la función eigh de Numpy, con un épsilon de $1E-20$ y una cantidad máxima de iteración igual a 10.000 . El resultado se puede ver en la figura 2, donde se nota que nuestros valores son cercanos a los esperados ya que el ecm de los autovectores generados por nuestra función ronda 10^{-17} en la mayoría de los casos.

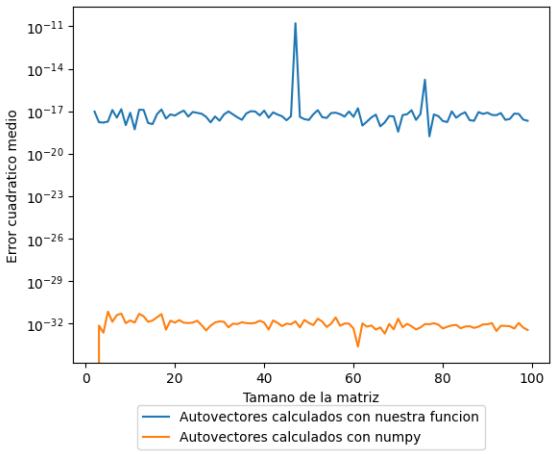


Figura 2: Comparación del error cuadrático medio resultante entre nuestra función y la de numpy

Luego, profundizamos en nuestra experimentación probando casos con matrices cuyos autovalores son cercanos. Con este objetivo, construimos una matriz de tamaño 10, cuyos 5 mayores autovalores son cercanos entre sí. En la figura 3 podemos ver como la cantidad de iteraciones realizadas para encontrar los primeros autovalores de la matriz utilizando el método de la potencia es mayor respecto de la cantidad necesaria con los otros autovalores.

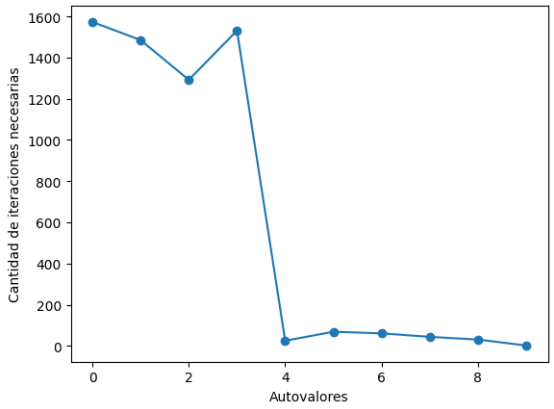


Figura 3: Cantidad de iteraciones realizadas por el método de la potencia para encontrar los autovalores en una matriz con autovalores cercanos

Sabemos que el método de la potencia no asegura la convergencia cuando existen dos autovalores iguales en la matriz. Para eso, realizamos una experimentación utilizando una matriz cuyos autovalores 4, 5 y 6 son iguales y observamos el error cuadrático resultante entre Av_i y $\lambda_i v_i$, siendo v_i el i-ésimo autovector de la matriz A y λ_i el autovalor correspondiente, esto lo calculamos para todos los autovectores y lo comparamos contra el de numpy. En la figura 4 podemos observar como, si bien nuestra función no da garantía de que funcione y nuestros errores son mas altos que los de Numpy, el error cuadrático medio es extremadamente pequeño.

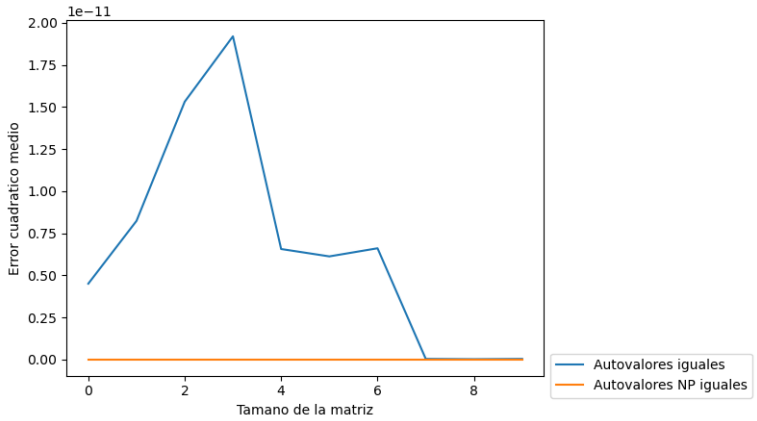


Figura 4: Error cuadrático medio para una matriz con autovalores iguales

3. Club de Karate

3.1. Desarrollo

La red de Club de Karate es una red de un club universitario que fue estudiada, en la que sucedió un conflicto y se dividió en dos partes, quienes eligieron el grupo del instructor y quienes eligieron el grupo del administrador. Queremos analizar la red, viendo la centralidad o influencia de los miembros, para poder predecir a que grupo opto cada miembro después del conflicto a partir de la estructura de la red. Además de calcular la centralidad de los nodos, nos gustaría saber cual es el autovector que mejor predice los grupos después del conflicto. Para esto, calculamos la matriz Laplaciana, calculamos todos sus autovectores y calculamos la correlación entre cada autovector y el vector que indica el grupo que viene con los datos. También queremos ver que tan bien estamos calculando nuestros autovectores, por lo que los comparamos con los autovectores que nos devuelve numpy, calculando el error cuadrático medio entre estos.

3.2. Resultados y discusión

Calculamos la centralidad del autovector asociado al mayor autovalor de la matriz de adyacencias de la red del Club de Karate, de esta forma podemos establecer una relación de orden entre los individuos de la red. Como se puede notar en la figura 5, hay dos nodos mas centrales, estos son el instructor y el administrador.

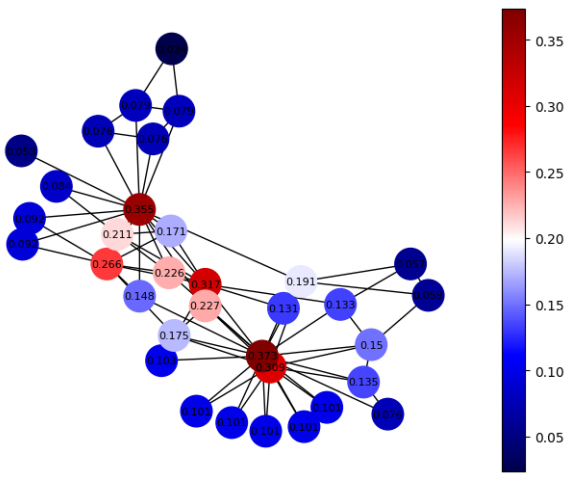


Figura 5: Centralidad del autovector de la red de karate

Analizando el gráfico podemos ver que centralidad tiene cada individuo en la red. Luego de calcular todos los autovalores y autovectores de la matriz Laplaciana, los queremos comparar con los autovalores y autovectores que nos devuelve numpy. Para esto, primero tenemos que ordenar los que recibimos de numpy, ya que nos los devuelve ordenados de menor a mayor autovalor y nosotros los queremos de mayor a menor. Luego calculamos el error cuadrático medio para cada autovector.

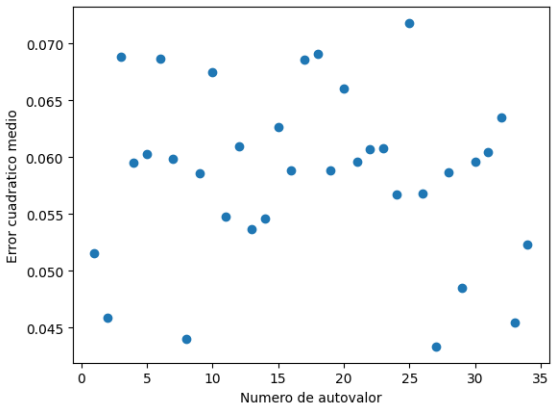


Figura 6: Error cuadrático medio entre los autovectores que calculamos y los que nos devuelve numpy

Para analizar que autovector predice mejor los grupos después del conflicto, calculamos la correlación entre nuestros autovectores y el vector que indica el grupo que viene con los datos, y buscamos el de mayor valor absoluto. Este valor es 0.814726339689374 y es la correlación del autovector que esta asociado al segundo autovalor mas chico.

4. Ego-Facebook

4.1. Desarrollo

Para analizar los datos de la red de ego-facebook queremos observar si es posible predecir el comportamiento de las amistades de la red basándonos en los atributos provistos. En primer lugar, definiremos un u umbral, cuyo valor definirá si dos nodos A y B son amigos. Si la cantidad de atributos compartidos por los nodos es mayor a u, entonces los nodos tienen una relación de amistad y por lo tanto están conectados. Luego, comparamos la matriz resultante con la matriz de relaciones original y vemos si efectivamente estamos pudiendo predecir correctamente el comportamiento de amistades. Utilizaremos PCA para mejorar la predicción de nuestros resultados.

4.2. Resultados y discusión

Calcular la matriz de similaridad como producto interno de la matriz de atributos nos dejará visualizar cuantos atributos comparten cada uno de los integrantes de la red con cualquier otro integrante de la red. Podemos observar en la figura 3 el coloreo de la matriz según la cantidad de atributos compartidos de los integrantes. Observar que la diagonal tiene valor 0 ya que es un valor no representativo.

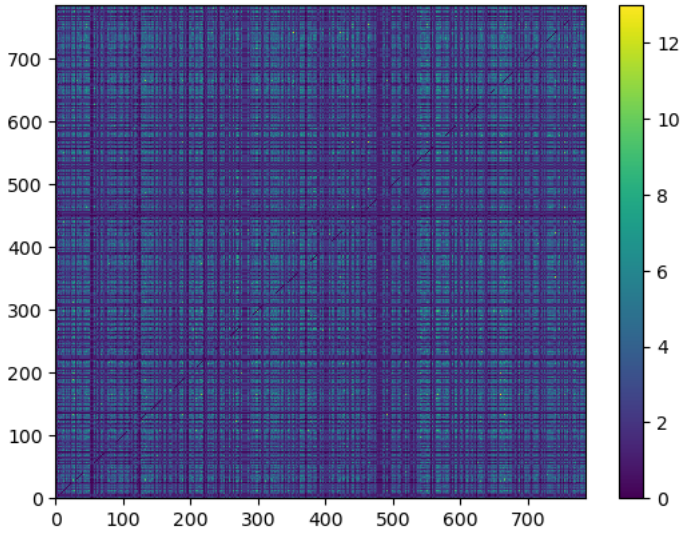


Figura 7: Matriz de similaridad con producto interno de la matriz de atributos

Luego, queremos encontrar una red que se asimile a la original en la medida de lo posible, prediciendo el resultado a partir de los atributos compartidos informados. Con ese objetivo, generamos para diferentes umbrales con valores entre 1 y 9 el grafo de amistades cuyos nodos son amigos si comparten más de u atributos. Luego, calculamos la correlación entre cada una de las matrices estiradas y la matriz de atributos original estirada.

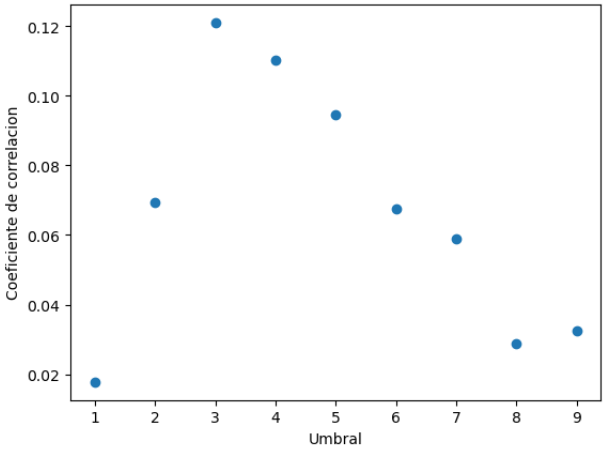


Figura 8: Variación del coeficiente de la correlación entre las matrices de adyacencia con diferentes umbrales con el método flatten

En la figura 8 se puede ver como varía el coeficiente de correlación de las matrices flatten para los diferentes grafos generados con diferentes umbrales. Allí podemos observar que alcanza su punto mas alto cuando el umbral es 3.
En la siguiente figura 9, medimos como varía el coeficiente de correlación de los autovalores de las matrices para los diferentes grafos generados con diferentes umbrales. Podemos observar que alcanza el punto máximo cuando el umbral es 7, llegando casi al 0.96.

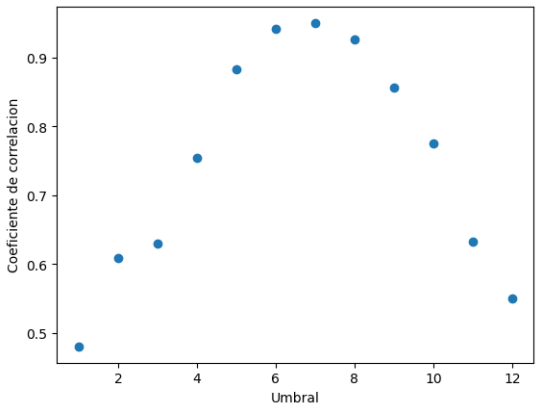


Figura 9: Variación del coeficiente de la correlación entre los autovalores de las matrices de adyacencia con diferentes umbrales

Podemos utilizar el método de análisis de componentes principales para mejorar el análisis de las redes, tomando diferentes k componentes principales y proyectándolos en un espacio para obtener una nueva matriz de atributos, con diferente dimensionalidad. En la figura 10 se puede ver como para diferentes k indicados en las etiquetas, varía el coeficiente de correlación de la matriz de atributos obtenida flatten con la original flatten, según el umbral utilizado. Podemos observar que el índice de correlación alcanza su punto más alto cuando k = 2, con un índice de correlación 0.13018731194207162 y un umbral 2.

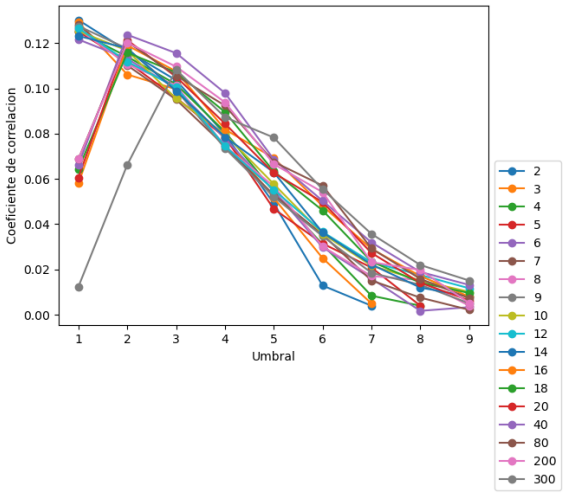


Figura 10: Variación del coeficiente de la correlación entre los autovalores de las matrices de adyacencia con diferentes umbrales

En la figura 11, podemos observar como para diferentes k indicados en las etiquetas, varía el coeficiente de correlación de los autovalores de matriz de atributos con los autovalores de la matriz original, según el umbral utilizado. Podemos observar que alcanza su punto más alto cuando k = 300 y el umbral es 600, con un coeficiente 0.9676613513858288.

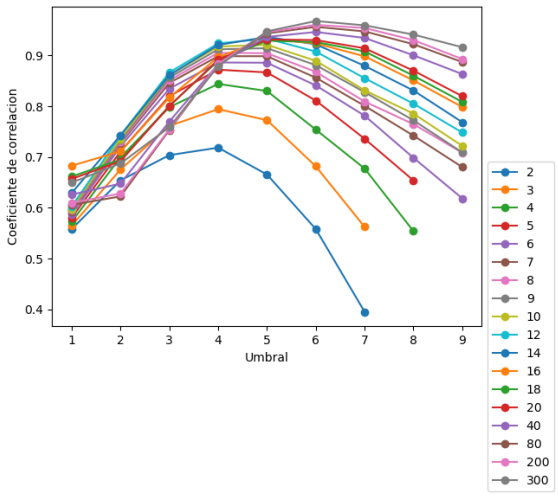


Figura 11: Variación del coeficiente de la correlación entre los autovalores de las matrices de adyacencia con diferentes umbrales

5. Conclusiones

A lo largo de este trabajo, hemos experimentado con diferentes métodos para estudiar redes sociales. Podemos observar que es posible aproximarse a una predicción de como actuarán estas redes sociales aplicando estudios con autovalores y autovectores, a la vez que es posible observar el comportamiento de la red. Son destacables los recursos que nos da trabajar con este tipo de técnicas para poder estudiar comportamientos sociales. Vemos la transversalidad de las aplicaciones de la ciencia y la matemática a lo largo de campos variados, y la importancia de seguir promoviendo el estudio y la investigación de los mismos.