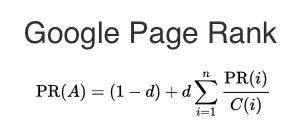
Raúl Jiménez Suárez

49235802R T2.

**PAGE RANK**

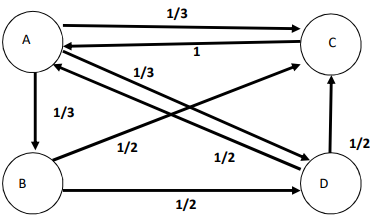
1. **Describe brevemente (no más de una carilla DIN-A4) en qué consiste el algoritmo. Explícalo con tus palabras y haz un esquema del proceso.**

El algoritmo PageRank es un algoritmo de Google lanzado en 1999 por sus fundadores Larry Page y Sergey Brin. Su función era medir la importancia y calidad de una página web en base a un criterio de autoridad que mide la cantidad y calidad de los retroenlaces que esta tenga, mientras más y mejores enlaces tenga, mayor puntuación tendrá. Su ecuación característica es la siguiente:



(diremos que el Pagerank de una página A es igual a 1 menos el damping factor (un valor constante que se usa como factor de reducción) más de nuevo el damping factor multiplicado por el sumatorio del pagerank recibido de otras páginas.)

Un ejemplo para ver el funcionamiento del algoritmo sería el siguiente:



Consideremos una web en la que solamente existen cuatro páginas. Cada una está representada por un nodo. La interpretación del mismo nos indica que según el algoritmo la página A será la primera en el ranking por su relevancia a la búsqueda, seguida por la C. Quedando las páginas D y B en tercer y cuarto puesto respectivamente.

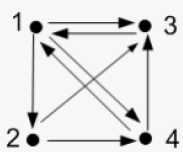
1. **¿Cuál es la hipótesis del surfista aleatorio? ¿En qué influye dicha hipótesis en el algoritmo Page Rank?**

También conocido como modelo de navegación aleatoria, proporciona una base para calcular el algoritmo de PageRank.

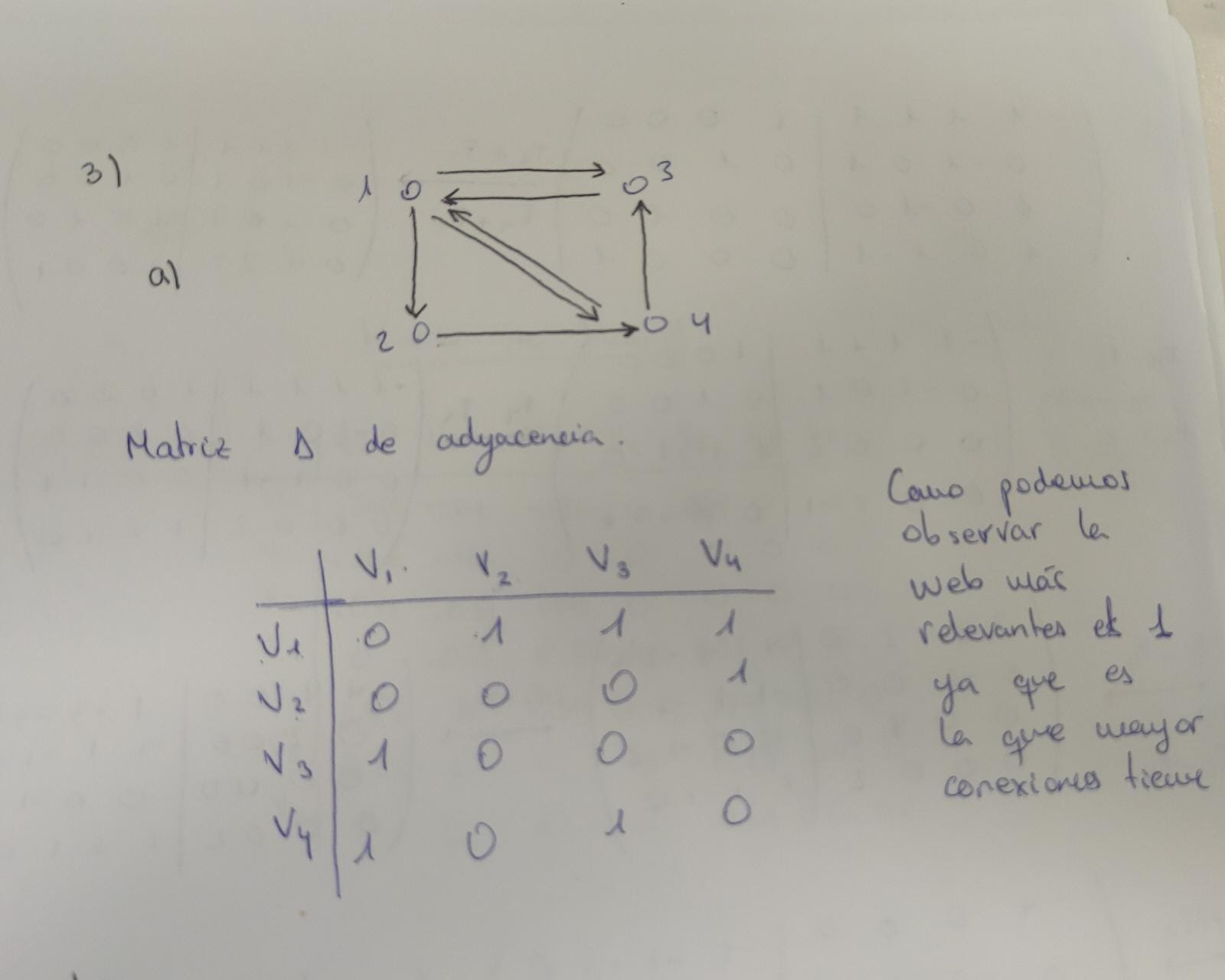
Un internauta se mueve a través de internet de dos maneras, introduciendo la URL o utilizando marcadores. En este modelo se asume que el enlace al que se dirige es completamente aleatorio, la probabilidad de que este en una página en concreto puede deducirse del PageRank (la probabilidad de que visite una página es la suma de todas las veces que ha accedido).

Por lo tanto las páginas web que a menudo están enlazadas también se visitan con frecuencia y tienen alto PageRank. Pero este valor sigue siendo reducido por un factor d. La razón es que un surfista aleatorio no siempre siguen indefinidamente un enlace, por lo que dependiendo de que la probabilidad de que un surfista que este siguiendo los enlaces no se rompa, d tiene un valor de 0 a 1 y cuanto más cerca este del 1, mayor será la probabilidad de que el enlace sea seguido.

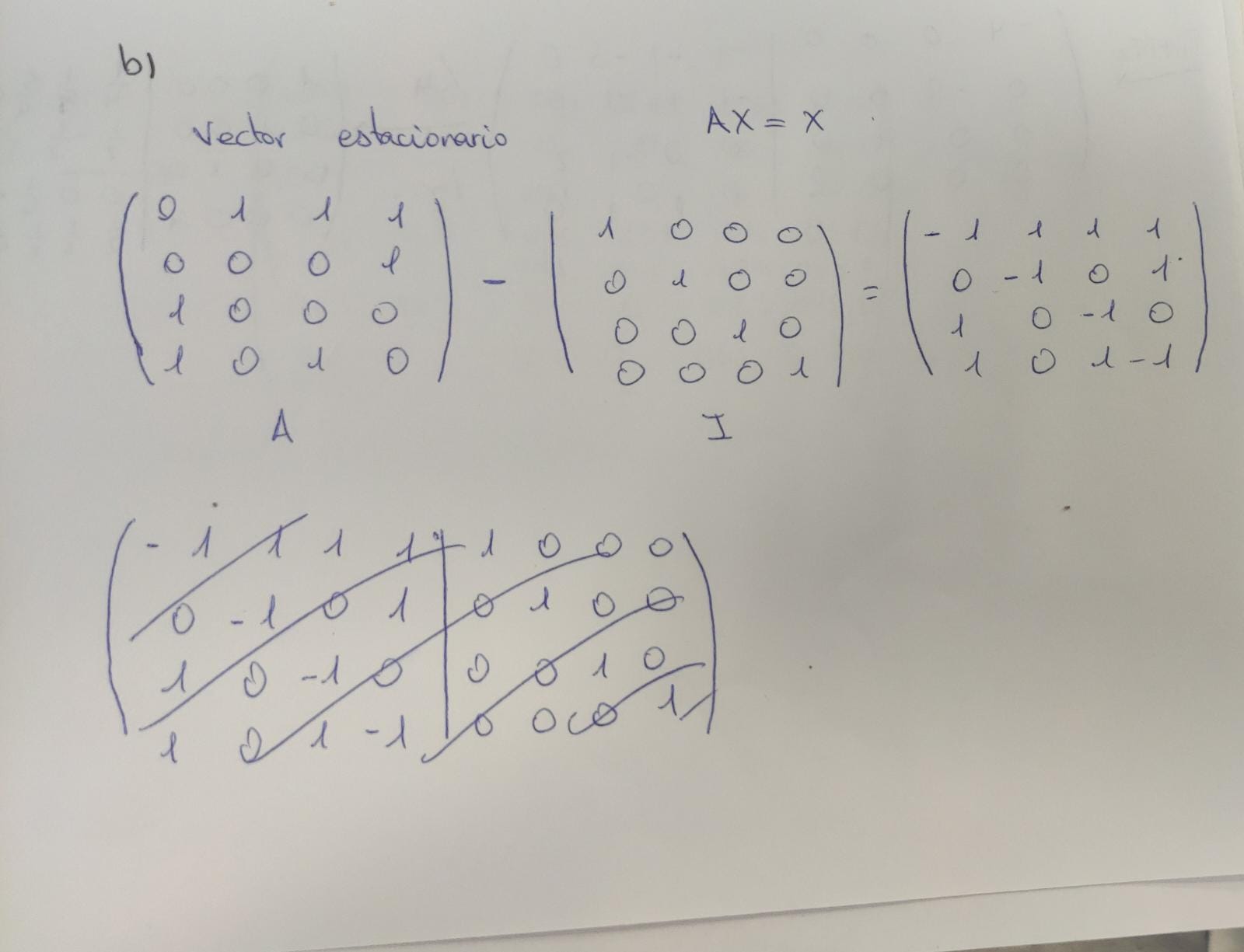
1. **Aplicación: Sea la siguiente composición indexada de páginas web:**

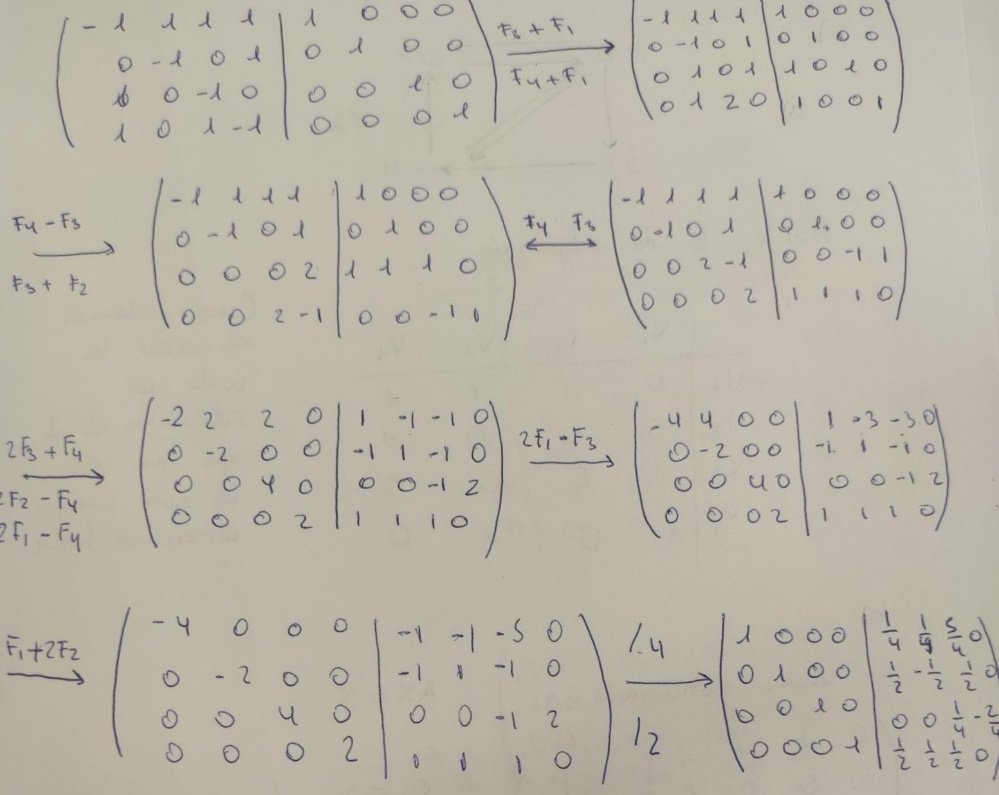
****

1. **Determina la matriz 𝑨 de adyacencia del grafo. ¿Qué página web es más relevante?**

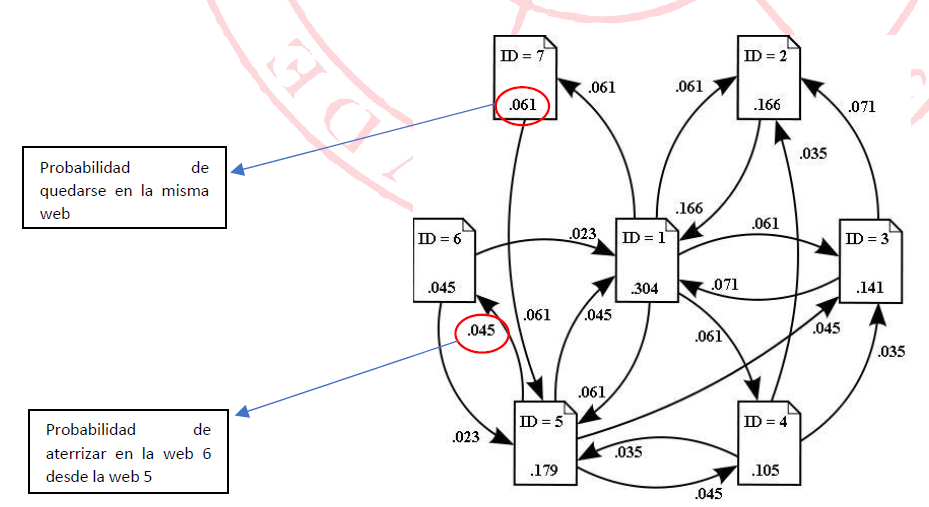
****

1. **Encuentra un vector estacionario, esto es, halla el vector 𝑋 tal que: 𝐴 𝑋=𝑋. ¿Qué interpretación tiene dicho vector estacionario en el algoritmo Page Rank?**

****

****

1. **Si consideramos el siguiente diagrama de flujo con 7 webs, donde los pesos en las aristas son las probabilidades de landing (aterrizar en una página web distinta al navegar desde otra). Determina la web más influyente y el**

**vector estacionario**

CODIGO EN MATLAB

A=[0.304 0.061 0.061 0.061 0.061 0 0; 0.166 0.166 0 0 0 0 0; 0.071 0.071 0.141 0 0 0 0; 0 0.035 0.035 0.105 0.035 0 0; 0.045 0 0.045 0.045 0.179 0.045 0; 0.023 0 0 0 0.023 0.45 0; 0 0 0 0 0.061 0 0.061]

%Como podemos observar web más influyente es la 1, ya que es la que mayor

%conexiones tiene y la segunda más influyente seria la 5.

% A\*X=X -> A\*X-X=I -> (A-I)^-1 \* (A-I)\*X = (A-I)^-1 \* I -> X = 8A-I)^-1 \* I

I=eye(7);

B=A-I

X=inv(B)

1. **Bibliografía: Haz mención a las referencias (artículos, libros, vídeos, webs,…) para documentar el trabajo.**

Juan Manuel Barriola y Milena Dotta (2016). ¿CÓMO FUNCIONA GOOGLE? EL ALGORITMO PAGERANK, DIAGRAMAS DE GRAFOS Y CADENAS DE MARKOV. Facultad de Ciencias Económicas, UBA.

Fernando Sancho Caparrini (2020). PageRank y el Surfista Aleatorio. De http://www.cs.us.es/~fsancho/?e=216.