# Práctica 1. El algoritmo PageRank

En esta práctica, repasaremos algunos comandos de Matlab, experimentaremos con distintos tipos de matrices de transición y veremos un ejemplo de aplicación de las cadenas de Markov: *PageRank*, el algoritmo de clasificación de páginas web que dio origen al popular buscador Google.

#### Vectores, matrices y ecuaciones lineales en Matlab

En este apartado se repasan todos los comandos de Matlab útiles en esta práctica. En el fichero **comandos\_matlab\_1.m** encontrará ejemplos de uso de cada comando. Ejecute aquellos que no conozca y compruebe su funcionamiento.

#### Probabilidades en n saltos. Ecuación de Chapman-Kolmogorov

Ejecute el fichero **matices.m**. Este fichero crea tres matrices: Pa, Pb y Pc. Evalúe a qué valores convergen las probabilidades de transición en *n* etapas, y a partir de los resultados, clasifique las clases contenidas en cada matriz.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pa | Pb | Pc |
| Número de estados transitorios |  |  |  |
| Número de clases recurrentes aperiódicas |  |  |  |
| Número de clases recurrentes periódicas |  |  |  |
| Es ergódica (si/no) |  |  |  |

#### Solución de las ecuaciones de balance

Programe la función **v = SolveErgodicDTMC(P).** El argumento (P) es una matriz de transición con una sola clase recurrente y aperiódica y el resultado (v) es el vector de estado estacionario de P.

#### Algoritmo PageRank

Programe la función **[r i] = PageRank(A, *α*)**, que devuelve el vector **r** con los *Ranks* y el vector **i** con los identificadores de las páginas, ordenados de mayor a menor *Rank*. El argumento ***α*** ≤ 1 es un escalar positivo y **A** es la matriz de incidencias definida así:

### Anexo: Fundametos de PageRank

El objetivo de *PageRank* es ordenar por importancia las direcciones web. La importancia de una página *j* se basa en el número de páginas que la referencian (páginas que contienen hipervínculos a *j*) así como la importancia de las páginas que la referencian. Supongamos que una página *i* referencia a *j*. La contribución de *i* en la importancia de *j* es proporcional a la importancia de *i* e inversamente proporcional al número total de hipervínculos de *i.*

Denominemos *H*(*j*) al conjunto de páginas que contienen hipervínculos a *j*. La importancia (*Rank*) de *j* () se define inicialmente como:

Donde es el número de hipervínculos que contiene la página k.

Si definimos la matriz de hipervínculos de la siguiente forma:

Resulta que es una matriz de transición y el ranking de las páginas se puede obtener resolviendo las ecuaciones de balance:

con , donde .

Sin embargo nada nos garantiza que sea ergódica. De hecho en la realidad no lo es. Por tanto, para resolver el problema del *PageRank* se redefine el *Rank* de la siguiente forma:

Donde es el número total de páginas y es un factor de ponderación ajustado a 0.85 generalmente. Es decir el *Rank* de *j* () es una media ponderada entre una importancia homogénea entre todas las páginas y el *Rank* definido anteriormente.

Para obtener la expresión matricial de *PageRank* definimos la matriz :

Donde es una matriz de N×N con todos sus elementos iguales a 1. Las ecuaciones quedan así:

con