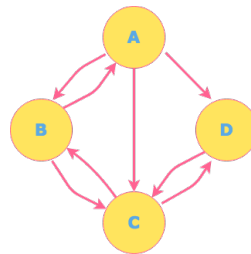


CADENAS DE MARKOV

CASO DE ESTUDIO: LA CIENCIA DETRÁS DEL ALGORITMO DE GOOGLE

El algoritmo PageRank, publicado por Brin & Page en 1998, es una técnica para clasificar páginas web según la importancia dada por los enlaces que llevan a dicha página. Intuitivamente, se considera internet como una red de “estados” o páginas web, a las que una persona llega mediante enlaces desde otras páginas (“navegando”). Supóngase que esta persona no tiene nada que hacer excepto pasar todo el tiempo en internet, entonces se detiene en una página, da clic en un enlace y va a otra página, y así continúa sucesivamente su navegación. Entonces, la siguiente página que visita depende de los “enlaces de salida” de la página actual. Para ilustrar esto, supóngase que internet está compuesto de solo 4 páginas web con los enlaces a otras páginas ilustrado a continuación:



Bajo estas condiciones, la página A tiene tres enlaces de salida hacia B, C y D. Su vector de conectividad sería:

$$L_A = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Pues A no tiene enlaces a sí mismo. Este vector puede convertirse en un vector de probabilidad, pues un navegante en la página A tiene 3 opciones posibles para ir a otra página y, dado que no sabemos cuál elegirá, cada una tiene $1/3$ de probabilidad de ser elegida, en otras palabras, el vector se divide para la suma de sus enlaces, y el vector queda:

$$L_A = \begin{pmatrix} 0 \\ 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$$

La construcción de estos vectores para las demás páginas web permite elaborar una matriz de conectividad y una matriz de probabilidad de cambio de página de todo el internet:

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \rightarrow L = \begin{pmatrix} 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/2 & 0 \\ 1/3 & 1/2 & 0 & 1 \\ 1/3 & 0 & 1/2 & 0 \end{pmatrix}$$

Si un navegante se encuentra en la página B en el tiempo t , ¿cuáles son sus opciones conforme empieza a clicar las páginas? Esto se resuelve mediante multiplicación de la matriz L con un vector posición en B, como se indica:

