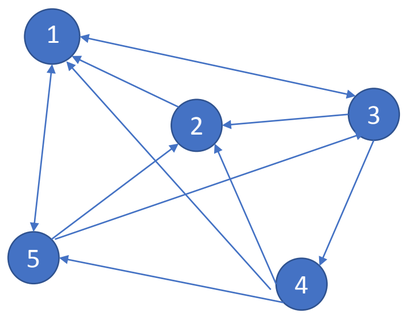
¿COMO FUNCIONA UN BUSCADOR DE INTERNET?

Jorge. Endelli, Claudia. Batallán, Fabiana. Sjoerdstra.

Universidad Nacional de Avellaneda

Tópicos: Tecnología, Enseñanza.

Imaginemos cada página de internet como un nodo del siguiente gráfico



*Con la información del grafo, construimos una matriz llamada "**Matriz de Incidencia del Grafo**" de la siguiente forma:*

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

En cada fila se observa la cantidad de páginas que salen de cada nodo. En cada columna se observa la cantidad de páginas que entran a cada nodo.

¿Cuándo una página es importante?

- Cuando una página tiene muchas citas.
- Cuando una página es citada por páginas importantes.

Postulado PageRank

"La importancia x_i de la página p_i es directamente proporcional a la suma de las importancias de las páginas que enlazan con ella".

Matricialmente tenemos:

$$M^tX^t = \lambda X^t$$

Teorema

Si M^t es la matriz de incidencia del grafo de internet y x = $(x_1;\ldots;x_n)$ donde cada $x_n \in R_{\geq 0}$ el vector de importancias,
entonces vale $M^tX^t = \lambda X^t$
donde $\lambda \in R_{\geq 0}$ es la constante de proporcionalidad.

Observación: El vector de importancias de las páginas web es un autovector (positivo) de la matriz M^t, y la constante de proporcionalidad λ es el autovalor asociado a este vector.

Teorema de Perron.
Sea M una matriz cuadrada con todos sus coeficientes positivos.
Entonces, existe un valor propio simple $\lambda > 0$ tal que $M^tv^t = \lambda v^t$,
donde v es el vector correspondiente, y tiene todas sus coordenadas positivas.
Este valor propio es mayor, en módulo, que todos los demás valores propios de M.
Cualquier otro vector propio positivo de M es un múltiplo de v.

Dado que nuestras matrices tienen muchos ceros, necesitamos un teorema mas fuerte que garantice cierta unicidad para matrices no negativas. Tenemos el siguiente teorema:

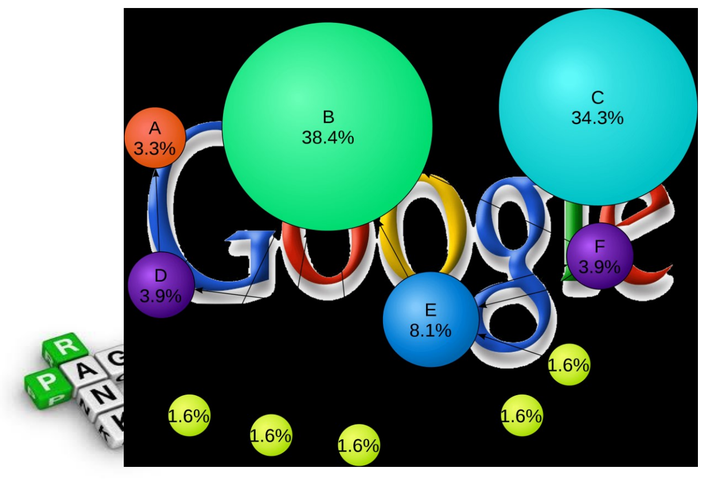
Teorema de Frobenius
Sea M una matriz cuadrada con entradas no negativas.
Si M es irreducible, entonces existe un valor propio simple $\lambda > 0$ tal que $M^tv^t = \lambda v^t$,
donde v es el vector correspondiente, y tiene todas sus coordenadas positivas.
Este valor propio es mayor o igual, en módulo, que todos los demás valores propios de M.
cualquier otro vector propio con entradas no negativas M es un múltiplo de v.

Todo parece estar resuelto para una matriz pequeña, pero que ocurre si nuestra matriz es enorme? Necesitamos una solución computacional y para eso, Google utiliza el método de las potencias.

Método de las potencias
Si una matriz cuadrada M es diagonalizable y tiene todos sus autovectores v_1 ;…; v_n numerados de tal manera que los autovalores correspondientes cumplan lo siguiente:
$\lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \ldots |\lambda_n|$ partiendo de $v_0 > 0$ tal que $v_0 = \alpha_1 v_1 + \ldots + \alpha_1 v_1$ con $\alpha_1 \neq 0$ entonces se tendrá $M^k v_0 = \alpha_1 {\lambda_1}^k v_1 + \ldots + \alpha_n {\lambda_n}^k v_n$ luego,

$$\lim_{k\to\infty} \frac{M^k v_0}{{\lambda_1}^k} = \alpha_1 v_1$$

Un múltiplo no trivial del autovector buscado.
Este método es el que utiliza Google para ordenar sus páginas con resultados bastante satisfactorios.



IVJIMA 2019
JORNADA DE INGENIERIA EN MATERIALES

Facultad de Ingeniería Universidad Nacional de La Plata
Universidad Nacional de La Plata
Departamento de Materiales
INMA
UNDAV Universidad Nacional de Avellaneda