Universidad Central de Venezuela

Facultad de Ciencias

Escuela de Computación

Asignatura: Cálculo Científico (6105)

Estudiante: Naranjo Sthory Alexanyer Antonio

Cédula de identidad: V – 26.498.600

**Tarea 6: El autovector de 458 billones de dólares (hasta ahora)**

Para el desarrollo de la solución al problema planteado en el enunciado del ejercicio, se optó por investigar, trabajar e implementar el algoritmo ***SpageRank*** desarrollado por los fundadores de Google, *Larry Page* y *Serguéi Brin*. Como siempre, se adjunta además del presente informe, una carpeta con el nombre ***Codes*** donde se almacenan los códigos fuentes escritos bajo el lenguaje de programación Matlab/Octave y una carpeta con el nombre ***Images*** con las imágenes de las gráficas que se hayan generado para sustentar lo explicado en cada sección del actual documento.

***Algoritmo SpageRank:* ¿Qué es?**

Una de las razones por las que *Google TM* es un motor de búsqueda tan eficaz es el algoritmo *PageRank TM* desarrollado por los fundadores de Google, *Larry Page* y *Serguéi Brin*, cuando eran estudiantes de posgrado en la Universidad de Stanford. El PageRank se determina enteramente por la estructura de enlaces de la ***World Wide Web*** (www). Se vuelve a calcular una vez al mes y no tiene en cuenta el contenido real de las páginas web ni las o consultas individuales. Entonces, para cualquier consulta concreta, Google encuentra las páginas de la que coinciden con esa consulta y las enumera en el orden de su PageRank.

Imagínese que navega por la web y pasa de una página a otra eligiendo al azar un enlace saliente de una página para llegar a la siguiente. Esto puede llevar a callejones sin salida en páginas sin enlaces salientes, o ciclos alrededor de camarillas de páginas interconectadas. Así que, una cierta fracción de las veces, simplemente se elige una página al azar de la Web. Este paseo aleatorio teórico se conoce como ***cadena de Markov*** o ***proceso de Markov***. La probabilidad límite de que un navegante aleatorio infinitamente dedicado visite una página es su PageRank. Una página tiene un alto rango si otras páginas con alto rango la enlazan a ella.

***Algoritmo SpageRank:* Análisis Teórico**

Sea el conjunto de páginas web a las que se puede llegar siguiendo una cadena de hipervínculos que comienza en una página raíz, y que es el número de páginas en . Para Google, el conjunto varía en realidad con el tiempo, pero en junio de 2004, superaba los .

Sea la matriz de conectividad de una parte de la web, es decir, si hay un hipervínculo a la página desde la página y en caso contrario. La matriz puede ser enorme, pero es muy dispersa. Su *j-enésima* columna muestra los enlaces de la *j-enésima* página. El número de ***nonzeros*** (valor positivo o negativo, no igual a cero) en es el número total de hipervínculos en .

Sean y , las sumas de las filas y columnas de :

Las cantidades y son el grado de entrada y el grado de salida de la página . Sea la probabilidad de que el paseo aleatorio siga un enlace. Un valor típico es . Entonces es la probabilidad de que se elija una página arbitraria y es la probabilidad de que se elija una página aleatoria concreta. Sea la matriz de cuyos elementos son,

Obsérvese que A proviene de escalar la matriz de conectividad por sus sumas de columna. La j-enésima columna es la probabilidad de saltar de la página a las demás páginas de la web. Si la página es un callejón sin salida, es decir, no tiene enlaces de salida, entonces asignamos una probabilidad uniforme de a todos los elementos de su columna. La mayoría de los elementos de son iguales a , la probabilidad de saltar de una página a otra sin seguir un enlace. Si y , entonces

La matriz A es la matriz de probabilidad de transición de la cadena de Markov. Sus elementos están todos estrictamente entre cero y uno y las sumas de sus columnas son todas iguales a uno. Un importante resultado de la teoría de matrices conocido como *teorema de Perron-Frobenius* se aplica a estas matrices. Concluye que una solución no nula de la ecuación,

Existe y es único dentro de un factor de escala. Si este factor de escala se elige de forma que,

Entonces es el vector de estado de la cadena de Markov y es el PageRank de Google. Los elementos de son todos positivos y menores que uno.

Además, el vector es la solución del sistema lineal singular y homogéneo,

***Algoritmo SpageRank:* Código Fuente**

***Algoritmo SpageRank:* Caso de Prueba**