



CURSO: CMP 0575 - TÓPICOS 2 (DATA MINING)
COLEGIO: POLITÉCNICO
Semestre: Primer Semestre 2018/2019 NRC: 1068

Tarea 2: Ejercicio usando la técnica *PageRank*

Problema:

1. Un representante del departamento de marketing de una trasnacional desea incluir en su punto de análisis de la reunión, el nivel de importancia de 20 páginas Web dentro de una sección WWW seleccionada aleatoriamente. Para dar respuesta a esta necesidad, usted como ingeniero propone utilizar la técnica *PageRank* sobre la sección WWW seleccionada.

Nota: Como punto de partida, se constará con un grafo dirigido de 20 nodos (ejemplo reducido en la fig. 1), el cual incluye 7 trampas (4 *spider traps* y 3 *dead ends* o viceversa). Informáticamente, para simplificar el punto de partida, el grafo se tradujo a una matriz de adyacencia donde cada escaque representa la cantidad de *in-links* al nodo activo (ver ejemplo de matriz).

Nodos	N1	N2	...	N50
N1	1	1	...	1
N2	0	1	...	0
...
N50	0	0	...	0

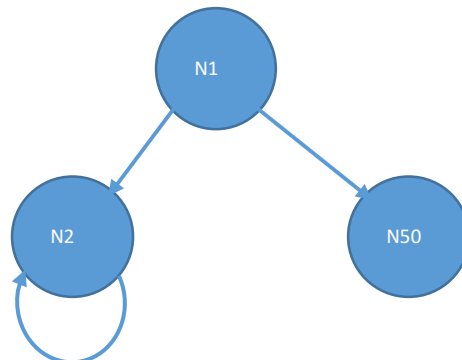


Fig. 1 Ejemplo reducido del grafo de 20 nodos, conteniendo 1 *spider trap* y 1 *dead end*.

Requisitos de evaluación:

- Es obligatorio el uso de la filosofía *PageRank* sobre un grafo dirigido.
- El algoritmo *random walk* debe incursionar de forma aleatoria (selección de caminos).
- Es obligatorio presentar la trazabilidad del vector R sobre el tiempo $t, t+1, t+2, \dots, t+n$; hasta que n sea el estado donde el algoritmo *power iterator* haya explorado todos los nodos del grafo (pueden presentar una matriz donde se vaya actualizando e incrementando dinámicamente el vector R por cada momento t).
- Es obligatorio que en la trazabilidad del vector R se pongan de manifiesto en las trampas y el *teleport* del *power iterator* para continuar con su desempeño.
- Cargar al D2L los códigos implementados (archivo compactado que incluye el ejecutable ej: el .JAR de java) dentro del plazo de entrega.